الإقتصاد القياسي

الجزء الأول

تأليف

د. جوجارات

D. Gujaratr

تعريب ومراجعة

أ. م. د. هند عبد الغفار عودة   أ. د. عفاف علي حسن الدش



الاقتصاد القياسي

الاقتصاد القياسي

الجزء الأول

تأليف

دامودار جيجاراتي
Damodar N. Gujaratic

ترجمة ومراجعة

أ.م.د. هند عبد الغفار عودة
رئيس قسم الإحصاء التطبيقي
كلية التجارة - جامعة حلوان

بالاشتراك مع
أ. د. عفاف علي حسين الدش
استاذة الإحصاء التطبيقي
كلية التجارة - جامعة حلوان



المملكة العربية السعودية - الرياض - هاتف: 4658523 - 4647531 + (009661)
ص. ب : 10720 - الرمز البريدي : 11443 - فاكس : 4657939 + (009661)

الطبعة الإنجليزية :

BASIC ECONOMETRICS

BY: Damodar N. Gudjratic

ردمك: 6 - 673 - 24 - 9960

© دار المريخ للنشر

المملكة العربية السعودية، الرياض، 1436هـ/2015م

المملكة العربية السعودية – الرياض – ص. ب : 10720 – الرمز البريدي : 11443

هاتف : 4647531 / 4658523 فاكس: 4657939 + (009661)

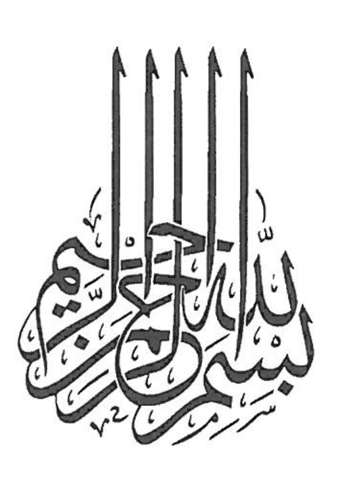
البريد الإلكتروني : Em.il: mrs@mrspubl.com

التوزيع داخل جمهورية مصر العربية والسودان وشمال افريقيا :

دار المريخ للنشر بالقاهرة – 4 شارع الفرات – المهندسين – الجيزة – الرمز البريدي: 12411

هاتف : 33376579 / 37609971 فاكس: 37609457 + (00202)

البريد الإلكتروني : Em.il: mrspub2002@Yhoo.com


بسم الله الرحمن الرحيم

محتويات الكتاب

(الجزء الأول : من الفصل الأول إلى الفصل الثالث عشر)
(الجزء الثاني : من الفصل الرابع عشر إلى الفصل الثاني والعشرين)

الموضـــوع — الصفحة

الجزء الأول

الفصل الأول

نماذج انحدار معادلة المتغير الواحد

الفصل الثاني

تحليل انحدار المتغير.. بعض الأفكار الأساسية

الفصل الثالث

نموذج انحدار متغيرين.. مشكلة التقدير

الفصل الرابع

نموذج الانحدار الخطي الطبيعي التقليدي

الفصل الخامس

انحدار متغيرين: تقدير الفترة.. اختبارات الفروض

الفصل السادس

توسيع نطاق نماذج الانحدار الخطية ثنائية المتغيرات

الفصل السابع

تحليل الانحدار المتعدد.. مشكلة التقدير

الفصل الثامن

تحليل الانحدار المتعدد.. مشكلة الاستدلال

الجزء الثاني

تحرير (تخفيف) فروض النموذج التقليدي

الفصل التاسع

نماذج الانحدار ذات المتغير الوهمي

الفصل العاشر

تعدد العلاقات الخطية..

ماذا يحدث إذا كانت المتغيرات المنحدرة مرتبطة؟

الفصل الحادي عشر

اختلاف التباين .. ماذا يحدث إذا كان تباين الخطأ غير ثابت؟

الفصل الثاني عشر

الارتباط الذاتي.. ماذا يحدث إذا كانت حدود الخطأ مرتبطة؟

الفصل الثالث عشر

نمذجة الاقتصاد القياسي: توصيف النموذج واختبارات التشخيص

الجزء الثالث

موضوعات في الاقتصاد القياسي

الفصل الرابـع عشر

نماذج الانحدار غير الخطية

الفصل الخامس عشر

نماذج الانحدار ذات الاستجابة النوعية

الفصل السادس عشر

نماذج انحدار البيانات

الفصل السابع عشر

نماذج الاقتصاد القياسي الديناميكية

نماذج الانحدار الذاتي.. ونماذج القيم الموزعة متأخراً

الجزء الثالث

نماذج المعادلات الآنية

الفصل الثامن عشر

نماذج المعادلات الآنية

الفصل التاسع عشر

مشكلة التوصيف

الفصل العشرون

طرق المعادلات الآنية

الفصل الحادي والعشرون

السلاسل الزمنية في الاقتصاد القياسي.. بعض المفاهيم الأساسية

الفصل الثاني والعشرون

السلاسل الزمنية في الاقتصاد القياسي: التنبؤ

ملحق A

مراجعة على بعض المفاهيم الإحصائية

ملحق B

مبادئ جبر المصفوفات

ملحق C

طريقة المصفوفات لنماذج الانحدار الخطي

ملحق D

مقدمة الكتاب

1 - ماهو الاقتصاد القياسي؟: WHAT IS ECONOMETRICS

التفسير الحرفي لمعنى الاقتصاد القياسي هو القياس الاقتصادي " economic measurement ". وبالرغم من أن القياس هو العنصر الرئيسي في الاقتصاد القياسي ، فإن الانتشار الواسع للاقتصاد القياسي ، يجعلنا لانغفل الأخذ في الاعتبار التعريفات التالية :

● **الاقتصاد القياسي** : هو نتيجة لرؤية معينة تجريبية لدور الاقتصاد role of economics بحيث تحتوي على تطبيق للإحصاء الرياضي mathematical statistics ، وللبيانات الاقتصادية التي تؤدي إلى مساندة تجريبية empirical support للنماذج المبينة باستخدام الاقتصاد الرياضي mathematical economics ، والحصول على نتائج عددية numerical results[1] .

● **كذلك يمكن تعريف الاقتصاد القياسي** : بأنه تحليل كمي quantitive analysis للظواهر الاقتصادية الفعلية . هذا التحليل مبني على النمو المتزامن concurrent development للنظرية والمشاهدة المرتبط بالطرق الملائمة للاستدلال[2] .

● **الاقتصاد القياسي** : يمكن تعريفه أيضاً بأنه أحد العلوم الاجتماعية social sciences الذي يتضمن أدوات النظرية الاقتصادية ، الرياضيات ، والاستدلال الإحصائي statistical inference للظواهر الاقتصادية[3] .

---

(1) Gerhard Tintner, Methodology of Mathematical Economics and Econometrics, The University of Chicago Press, Chicago, 1968, p. 74.

(2) P. A. Samuelson, T. C. Koopmans, and J. R. N. Stone, "Report of the Evaluative Committee for Econometrica," Econometrica, vol. 22, no. 2, April 1954, pp. 141–146.

(3) Arthur S. Goldberger, Econometric Theory, John Wiley & Sons, New York, 1964, p. 1.

● الاقتصاد القياسي : هو العلم الذي يهتم بالتحديد التجريبي the empirical determination للقوانين الاقتصادية[4]. وتظهر براعة المتخصصين في الاقتصاد القياسي econometricians في وضع الفروض assumptions بمواصفات كافية sufficiently realistic أيضاً، بحيث يتيح لهم تحقيق أفضل المزايا الممكنة من خلال البيانات المتاحة[5].

فهم يمثلون إضافة إيجابية حقيقية في محاولة إزالة الرؤية العامة السطحية poor public image للاقتصاد (كمي أو غير ذلك) . حيث مازالت كثير من النقاط تختلف فيها آراء الاقتصاديين[6] .

وتهدف أساليب البحث في الاقتصاد القياسي أساساً إلى استخدام أساليب الاستدلال الإحصائي للربط بين النظرية الاقتصادية والقياسات الفعلية ، حيث تعتبر أساليب الاستدلال الإحصائي بمثابة جسر يربط بين النظرية الاقتصادية والقياسات الفعلية[7] .

2 - لماذا العرض المحايد؟ : WHY ASEPARATE DISCIPLINE

من التعريفات السابق تقديمها ، يتضح أن الاقتصاد القياسي هو توليفة مترابطة من : النظرية الاقتصادية ، والاقتصاد الرياضي ، والإحصاء الاقتصادي ، والإحصاء الرياضي . ومع ذلك ، فإن الموضوع يستحق الدراسة للأسباب التالية :

● النظرية الاقتصادية تحديدات statements أو فروض hypotheses في معظم الأحيان تكون وصفية qualitative .

فعلى سبيل المثال ، نظرية الاقتصاد الجزئي microeconomic تقرر states أن انخفاض السعر في سلعة ما يتوقع أن يؤدي إلى زيادة الكمية المطلوبة من السلعة في حالة ثبات باقي العوامل الأخرى . هكذا تفترض النظرية الاقتصادية العلاقة العكسية بين السلع، والكمية المطلوبة من سلعة ما . ولكن لم تمدنا النظرية بمقياس عددي numerical measure يقيس مقدار الزيادة في الكمية المطلوبة الناشئ من نقص معين في سعر السلعة . ويعتبر عمل المختص بالاقتصاد القياسي هو تحديد هذا المقياس

(4) H. Theil, Principles of Econometrics, John Wiley & Sons, New York, 1971, p. 1.

(5) E. Malinvaud, Statistical Methods of Econometrics, Rand McNally, Chicago, 1966, p. 514.

(6) Adrian C. Darnell and J. Lynne Evans, The Limits of Econometrics, Edward Elgar Publishing , Hants, England, 1990, p. 54.

(7) T. Haavelmo, "The Probability Approach in econometrics," Supplement to Econometrica, vol. 12, 1944, preface p. iii.

العددي ، أو بعبارة أخرى ، الاقتصاد القياسي يمدنا بالمحتوى التجريبي empirical content لمعظم النظرية الاقتصادية .

ويعتبر الاهتمام الرئيسي للاقتصاد الرياضي ، هو أن يعبر عن النظرية الاقتصادية في صياغة رياضية mathematical form (معادلات equations) بدون الأخذ في الاعتبار القدرة القياسية على المراجعة التجريبية empirical verification للنظرية .

وكما ذكرنا سابقاً ، أن الاقتصاد القياسي يهتم أساساً بالمراجعة التجريبية للنظرية الاقتصادية . وكما سوف نوضح فيما بعد ، أن المختص بالاقتصاد القياسي غالباً ما يتناول المعادلات الرياضية التي يقترحها المختص بالاقتصاد الرياضي ، ولكنه يضع هذه المعادلات في شكل يمكن من الاختبار التجريبي empirical testing . وتحويل المعادلات الرياضية إلى معادلات اقتصادية قياسية يتطلب براعة ومهارة عملية عالية .

ويهتم الإحصاء الاقتصادي أساساً economic statistics بجمع collecting ، وميكنة processing أو عرض البيانات الاقتصادية presenting economic data في أشكال بيانية وجداول . وبصفة عامة ، يكون اهتمام المختص بالإحصاء الاقتصادي هو تجميع البيانات مثل بيانات الناتج القومي الإجمالي gross national product ، العمالة employment أو البطالة unemployment ، الأسعار price ، . . الخ . حيث يتم تجميع البيانات بشكل دوري ومستمر . أما اختبار النظريات الاقتصادية من خلال البيانات ، يكون من اختصاص المختص بالاقتصاد القياسي ، وليس من اختصاص المختص بالإحصاء الاقتصادي .

وبالرغم من أن الإحصاء الرياضي يمدنا بالأدوات المستخدمة في الاقتصاد القياسي ، فإنه غالباً يحتاج إلى طرق خاصة special methods في مراجعة الطبيعة الفريدة unique nature لمعظم البيانات الاقتصادية . بمعنى أن البيانات يجب أن تكون فعلية وليست بيانات مولدة generated من تجارب تحكمية controlled experiment .

في الاقتصاد القياسي القائم ببناء النموذج mosleler غالباً يتعامل مع البيانات المشاهدة observational data وليست بيانات تجريبية experimental data . إن النمذجة التجريبية empirical modeling في الاقتصاد القياسي تتطلب توافر الآتي :

أولاً : أن يكون القائم ببناء النموذج على دراية بمهارات متعددة ومختلفة ، وليس مجرد محلل للبيانات التجريبية فقط .

ثانياً : ضرورة الفصل بين القائم بتجميع البيانات ، والقائم بتحليل البيانات ، كذلك أن يكون القائم ببناء النموذج ملماً إلماماً جيداً بالبيانات[8] .

3 - المنهجية في الاقتصاد القياسي:

METHODOLOGY OF ECONOMETRICS

كيف يتناول المتخصصون في الاقتصاد القياسي بالتحليل المشاكل الاقتصادية؟ أو بعبارة أخرى : ماهي المنهجية التي يتبعها المتخصصون في الاقتصاد القياسي في تحليل المشاكل الاقتصادية؟

بالرغم من وجود مدارس متعددة في وضع منهجية الاقتصاد القياسي ، فإننا سوف نقدم المنهجية التقليدية أو الكلاسيكية traditional or classical methodology ، حيث تعتبر هذه المنهجية هي السائدة في البحث التجريبي empirical research في الاقتصاد والعلوم الاجتماعية والسلوكية الأخرى[9] . وبصفة عامة ، فإن منهجية الاقتصاد القياسي التقليدية هي العمل على التوالي في المحاور التالية :

1 - توضيح الجانب النظري أو الفروض : Statement of theory or hypothesis

2 - تحديد النموذج الرياضي المتوافق مع الجانب النظري :
Specification of the mathematical model of the theory

3 - تحديد النموذج الإحصائي أو الاقتصاد القياسي :
Specification of the statistical, or econometric, model

4 - الحصول على البيانات المطلوبة : Obtaining the data

5 - تقدير معلمات نموذج الاقتصاد القياسي :
Estimation of the parameters of the econometric model

6 - اختبار الفروض : Hypothesis testing

7 - التنبؤ : Forecasting or prediction

---

(8) Aris Spanos, Probability Theory and Statistical Inference: Econometric Modeling with Observational Data, Cambridge University Press, United Kingdom, 1999, p. 21.

(9) For an enlightening, if advanced, discussion on econometric methodology, see David F. Hendry, Dynamic Econometrics, Oxford University Press, New York, 1995. See also Aris Spanos, op. cit.

8 – استخدام النموذج في المراقبة أو تحديد تأثير السياسة :

Using the model for control or policy purposes

ولتوضيح الخطوات السابقة ، سوف نوضح ذلك من خلال نظرية كينز للاستهلاك ، على النحو التالي :

1 - توضيح الجانب النظري أو الفروض: Statement of theory or hypothesis

قرر كينز أن القوانين الأساسية لعلم النفس تفيد أن الرجال (أو السيدات) يميلون إلى زيادة الاستهلاك عند زيادة الدخل ، بحيث لاتزيد زيادة الاستهلاك عن زيادة الدخل[10] .

وباختصار ، افترض كينز أن الميل الحدي للاستهلاك marginal propensity to consume (MPC) ، أي معدل تغير الاستهلاك الراجع إلى تغير الدخل بوحدة واحدة ، وهذا المعدل أكبر من الصفر وأقل من الواحد .

2 - تحديد النموذج الرياضي للاستهلاك :

Specification of the mathematical model of consumption

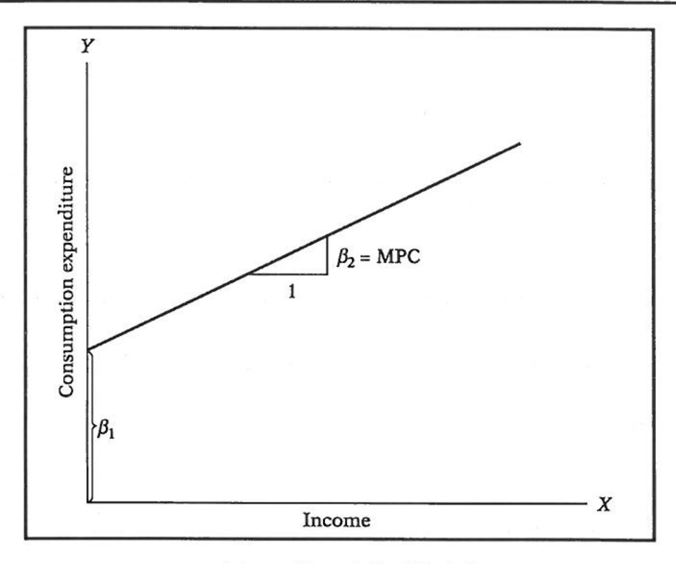
بالرغم من أن كينز افترض أن العلاقة موجبة positive relationship بين الاستهلاك consumption والدخل income ، لكنه لم يحدد الصياغة الدقيقة precise form للعلاقات الأساسية بين الاستهلاك والدخل .

وللتبسيط ، فإن المختص بالاقتصاد الرياضي ، يمكن أن يقترح الصياغة التالية للتعبير عن العلاقة بين الاستهلاك والدخل لكينز .

$$Y = \beta_1 + \beta_2 X \qquad 0 < \beta_2 < 1 \qquad \textbf{(I.3.1)}$$

حيث Y تساوي الإنفاق على الاستهلاك ، X تساوي الدخل ، كذلك معلمات النموذج β_1 ، β_2 يمثلان معلمات النموذج ، والشكل التالي ، يوضح بيانياً المعلمات β_1 ، β_2 على الترتيب .

(10) John Maynard Keynes, The General Theory of Employment, Interest and Money, Harcourt Brace Jovanovich, New York, 1936, p. 96.



شكل (I) دالة كينز للاستهلاك

ويسمى النموذج في (1.3.I) في الاقتصاد بدالة الاستهلاك consumption function . والنموذج ببساطة هو فئة المعادلات الرياضية mathematical equations . فإذا كان النموذج يشتمل على معادلة واحدة سمي بنموذج المعادلة الواحدة single-equation- model ، وفي حالة احتواء النموذج على أكثر من معادلة ، فإنه يسمى بنموذج متعدد المعادلات (1.3.I) ، حيث Y يمثل المتغير التابع dependent variable ، ويمثل X المتغير (أو المتغيرات) المستقلة أو المفسرة independent or explanatory variable (s) .

هكذا يعتبر الإنفاق الاستهلاكي هو المتغير التابع ، والدخل هو المتغير المستقل في دالة الاستهلاك الكنزية (1.3.I) .

3 - تحديد نموذج الاقتصاد القياسي للاستهلاك :

Specification of the econometric model of consumption

النموذج الرياضي المجرد لدالة الاستهلاك في (I.3.1) يمثل استفادة محدودة بالنسبة للمتخصص في الاقتصاد القياسي ، حيث إنه يفترض العلاقة الصحيحة أو اليقينية exact or deterministic relationship بين الاستهلاك والدخل ، وذلك يعني قيمة معينة واحدة للدخل X ، توجد قيمة واحدة للاستهلاك Y . ولكن العلاقة بين المتغيرات الاقتصادية بصفة عامة تكون غير يقينية undeterministic or inexact .

ولكن لو أننا أخذنا عينة مكونة من 50 أسرة أمريكية ، وتم تسجيل الدخل والإنفاق لكل أسرة . ثم تم رسم البيانات ، حيث يحدد المحور الأفقي X ليمثل الدخل ، والمحور الرأسي Y ليمثل الاستهلاك ، حيث تمثل النقطة الواحدة في المستوى دخل الأسرة واستهلاك نفس الأسرة . فنجد أن جميع النقاط لاتقع على خط واحد كما هو مفترض في العلاقة (1.3.I) . وذلك يرجع إلى أنه توجد متغيرات أخرى بالإضافة للدخل تؤثر على الاستهلاك مثل حجم الأسرة ، أعمار أفراد الأسرة ، ديانة الأسرة ، . . الخ . فكل هذه المتغيرات لا تؤثر على الاستهلاك ، ولإتاحة الأخذ في الاعتبار تأثير المتغيرات الأخرى (غير الدخل) ، فإن المختص بالاقتصاد القياسي يقوم بتعديل الدالة اليقينية في (1.3.I) إلى أخرى على النحو التالي :

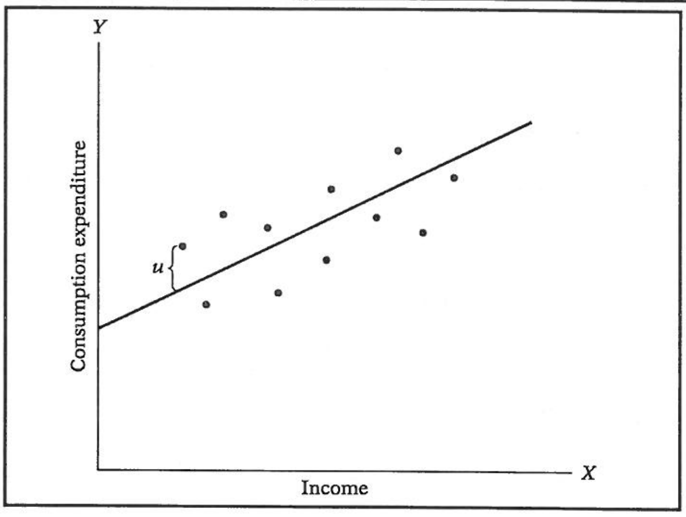
$$Y = \beta_1 + \beta_2 X + u \qquad \textbf{(2.3.I)}$$

حيث يتمثل تأثير المتغيرات الأخرى (غير الدخل) المؤثرة على الاستهلاك في إضافة المتغير u الذي يسمى بحد الخطأ أو التشويش disturbance or error . فحد الخطأ u هو في الحقيقة متغير عشوائي random variable ، له خصائص احتمالية probability properties . فالمتغير u يمثل تأثير مجموعة المتغيرات الأخرى (غير الدخل) المؤثرة على الاستهلاك ، ولكن لاتوجد في المعادلة بشكل صريح . والمعادلة (2.3.I) هي نموذج الاقتصاد القياسي في المثال محل الاعتبار . ويعتبر النموذج (2.3.I) نموذج انحدار خطي linear regression model .

وفي هذا الكتاب ، سوف نتناول بالتفصيل نموذج الانحدار الخطي .

والنموذج (2.3.I) ، يفترض العلاقة الخطية غير اليقينية بين الاستهلاك (Y) والدخل (X) ، حيث يأخذ في الاعتبار تأثير المتغيرات الأخرى المؤثرة بشكل إجمالي على الاستهلاك ، ويتمثل ذلك في إضافة المتغير العشوائي u .

ونموذج الاقتصاد القياسي (2.3.I) يوضح بيانياً في الشكل (2) على النحو التالي :



شكل (2) نموذج الاقتصاد القياسي لدالة كينز للاستهلاك

4 - الحصول على البيانات : Obtaining Data

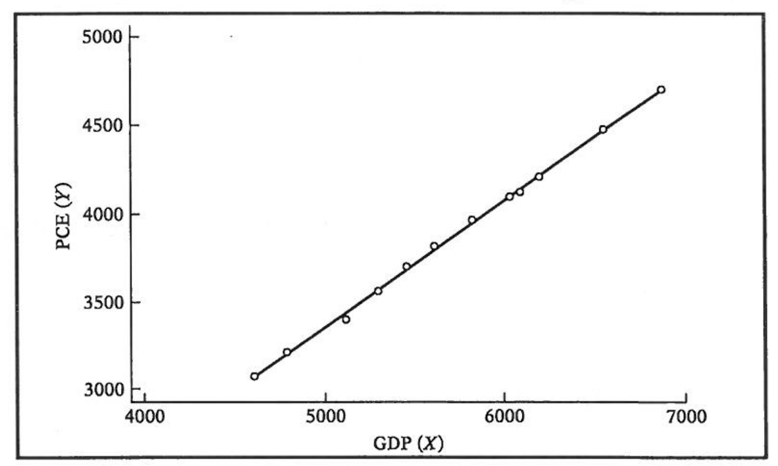
لتقدير معلمات النموذج (2.3.I) أي الحصول على القيم العددية numerical values لكل من β_1 و β_2 فإن ذلك يتطلب بيانات data . في الباب الثاني ، سوف نتناول بالتفصيل أهمية البيانات في التحليل الاقتصادي . أما بالنسبة لحساب كل من β_1 و β_2 في المثال محل الاعتبار ، سوف نعتبر البيانات في الجدول التالي جدول (1-1) .

جدول (I) بيانات عن Y (الإنفاق الاستهلاكي الشخصي) ، X (الناتج المحلي الإجمالي) في الفترة 1982–1996 بالبليون دولار بالنسبة لسنة 1992

	Y	X
1982	3081.5	4620.3
1983	3240.6	4803.7
1984	3407.6	5140.1
1985	3566.5	5323.5
1986	3708.7	5487.7
1987	3822.3	5649.5
1988	3972.7	5865.2
1989	4064.6	6062.0
1990	4132.2	6136.3
1991	4105.8	6079.4
1992	4219.8	6244.4
1993	4343.6	6389.6
1994	4486.0	6610.7
1995	4595.3	6742.1
1996	4714.1	6928.4

Source: Economic Report of the President, 1998, Table B–2, p. 282.

فالبيانات في الجدول مرتبطة باقتصاد U.S خلال الفترة 1981–1996 . فالمتغير Y بالجدول هو الإنفاق الاستهلاكي الشخصي (PCE) ، التجميعي aggregate بالنسبة للاقتصاد ككل ، والمتغير X هو الناتج المحلي الإجمالي (GDP) ، وكل من Y ، X مقاسة بالمليون دولار بالنسبة لسنة 1992 . فالبيانات بيانات حقيقية real منسوبة لأسعار سنة 1992 . وشكل (3) يوضح العلاقة بين (X) GDP و (Y) PCE



شكل (3) يوضح العلاقة بين الإنفاق الاستهلاكي الشخصي (PCE) والناتج المحلي الإجمالي X (GDP) خلال الفترة 1982–1996

5- تقدير نموذج الاقتصاد القياسي: Estimation of the Econometric model

وبعد الحصول على البيانات ، تصبح الخطوة التالية هي كيفية إيجاد القيم العددية لـ β_1 و β_2 من خلال البيانات . وسوف نناقش ذلك بالتفصيل في الباب الثالث . وهنا يتضح أن الأسلوب الإحصائي statistical technique لتحليل الانحدار regression analysis يعتبر الأداة الرئيسية في الحصول على قيم التقديرات β_1 و β_2 . فباستخدام هذا الأسلوب والبيانات في جدول (1) نحصل على قيم β_1 و β_2 على النحو التالي 0.7064 و 184.08– على الترتيب ، وبالتالي تصبح الدالة التقديرية للاستهلاك على النحو التالي :

$$\hat{Y} = -184.08 + 0.7064X_i \qquad (3.3.I)$$

حيث $\hat{Y}$ تشير إلى القيمة التقديرية لـ Y[11] ، ودالة الاستهلاك التقديرية (وتسمى بخط الانحدار regression line) كما هو موضح في شكل (3) يتضح أن خط الانحدار

(11) As a matter of convention, a hat over a variable or parameter indicates that it is an estimated value.

هو أفضل خط يتوسط النقط . ومن الشكل أيضًا نجد أن معامل الانحدار (أي الميل الحدي للاستهلاك MPC) خلال الفترة 1982–1996 تساوي تقريباً 0.70 . وهذا يعني أن زيادة الدخل الحقيقي بدولار واحد سوف تؤدي في المتوسط إلى زيادة المنفق على الاستهلاك[12] بـ 0.70 دولار أي 70 سنتًا . ويقال في المتوسط لأن العلاقة بين الاستهلاك والدخل علاقة غير يقينية underterministic or inexact ، كما هو موضح في شكل (3) حيث إن جميع النقط لاتقع على خط الانحدار .

وببساطة ، يمكن القول إن متوسط (أو توقع) الإنفاق على الاستهلاك يزيد بـ 70 سنتًا لزيادة الدخل بدولار واحد .

6 - اختبارات الفروض : Hypothesis testing

بافتراض أن النموذج الذي تم توقيعه يعتبر تقريبًا جيدًا للعلاقة الفعلية ، لذلك يجب تقديم معيار مناسب suitable criteria لتحديد أي التقديرات التي حصلنا عليها في المعادلة (3.3.I) يكون وفقاً للتوقعات في الجانب النظري الذي يتم اختياره .

ويرى الاقتصاديون الإيجابيون positive economists أمثال Milton ، Friedman أن الجانب النظري أو الفرضي إن لم يمكن إثباته بالرجوع إلى الدليل التجريبي ، فإنه لايرقى كجزء من البحث العلمي[13] . وكما ذكرنا سابقاً ، توقع كينز أن MPC يجب أن يكون موجبًا ولكن أقل من 1 ، ووجدنا في المثال أن MPC تساوي تقريباً 0.70 وهذا يتفق مع افتراض كينز أن قيمة MPC بين الصفر ، وواحد . ولكن قبل قبولنا بهذا الدليل ، يجب أن نتحقق من أن قيمة الـ MPC أقل من الواحد يرجع إلى صحة الفروض ، وذلك بإثبات إحصائياً أن MPC قيمته تقع بين الصفر والواحد ، كذلك التحقق من قيمة MPC تساوي 0.7 في المثال لايرجع إلى خاصية معينة في بيانات المثال .

وقبول الفروض النظرية أو رفضها في الاقتصاد يعتمد على عينة ملائمة تعتمد أساساً على جزء في النظرية الإحصائية Statistical theory يسمى الاستدلال الإحصائي statistical inference (اختبارات الفروض hypothesis testing) . وفي هذا

---

(12) Do not worry now about how these values were obtained. As we show in Chap. 3, the statistical method of least squares has produced these estimates. Also, for now do not worry about the negative value of the intercept.

(13) See Milton Friedman, "The Methodology of Positive Economics, "Essays in Positive Economics, University of Chicago Press, Chicago, 1953.

الكتاب ، سوف نوضح أن عملية الاستدلال inference process هي المرشد الفعلي لقبول أو رفض الفروض النظرية .

7 - **التوقع أو التنبؤ : Forecasting or prediction**

في حالة عدم رفض النموذج الذي تم اختياره ، فإنه يمكن استخدامه في إيجاد القيم التنبؤية للمتغير التابع dependent or forecast variable y ، عند قيم معينة للمتغير المستقل أو المفسر predictor variable X . ولتوضيح ذلك فلنعتبر النموذج (3.3.I) ونرغب في التنبؤ بمتوسط الإنفاق الاستهلاكي $\hat{Y}$ في سنة 1997 بافتراض أن قيمة GDP في سنة 1997 تساوي 7269.8 بليون دولار[14] في المعادلة (3.3.I) نحصل على :

$$\hat{Y}_{1997} = -184.0779 + 0.7064\,(7269.8)$$
$$= 4951.3167 \quad \approx 4951 \text{ بليون دولار} \qquad \textbf{(4.3.I)}$$

وهذا يعني أن متوسط الإنفاق الاستهلاكي في سنة 1997 يساوي تقريباً 4951 بليون دولار . في حين أن القيمة الفعلية في سنة 1997 تساوي 4913.5 بليون دولار ، أي أن :

$$Y_{1997} = 4913.5 \text{ بليون دولار}$$

وبالتالي ، فإننا نلاحظ وجود فرق بين القيمة المقدرة للإنفاق الاستهلاكي $\hat{Y}$ ، والقيمة الفعلية للإنفاق الاستهلاكي Y ، حيث تزيد القيمة المقدرة عن الفعلية بـ 37.8167 بليون دولار ، أي أن $\hat{Y}_{1997} - Y_{1997} = 4951.3167 - 4913.5 = 37.8167$

ويسمى الفرق $(\hat{Y}-Y)$ بخطأ التنبؤ forecast error ، وفي هذا المثال ، يمثل هذا الخطأ نسبة 0.76% من القيمة المقدرة لـ $\hat{Y}$ حيث :

$$\frac{37.8167}{4951.3167} \times 100 = \%0.76$$

وفي الفصول التالية عند مناقشة نماذج الانحدار الخطي ، سوف نتناول بالتفصيل الخطأ error في التنبؤ وأهميته . ويستخدم النموذج المقدر في (3.3.I) في اتخاذ القرار . فعلى سبيل المثال ، إذا قرر الرئيس اقتراح تخفيض ضرائب الدخل : ما تأثير هذه السياسة على الدخل ، الإنفاق الاستهلاكي ، وأخيراً على العمالة؟ .

(14) Data on PCE and GDF were available for 1997 but we purposely left them out to illustrate the topic discussed in this section. As we will discuss in subsequent chapters, it is a good idea to save a portion of the data to find out how well the fitted model predicts the out-of-sample observations.

وبافتراض حدوث زيادة في الإنفاق الاستثماري نتيجة للسياسة المقترحة . ماهو تأثير ذلك على الاقتصاد؟ بالإضافة إلى أن النظرية الاقتصادية الكلية macroeconomic theory توضح أن تغير الدخل سوف يؤدي إلى تغير في الإنفاق الاستثماري ، ويقاس هذا التغير بمعامل الدخل M income multiplier حيث :

$$M = \frac{1}{1 - \text{MPC}} \qquad (5.3.\text{I})$$

ففي نموذج (5.3.I) نجد أن MPC تساوي 0.70 ، وبالتالي نجد أن M تساوي 3.33 . وهذا يعني أن زيادة (أو نقص) الاستثمار بدولار ، سوف يؤدي إلى زيادة (أو نقص) الدخل بأكثر من 3 دولارات (قيمة M) .

ومن ثم تعتبر قيمة MPC هي القيمة الحرجة في تقدير M ، فنجد أن نموذج الانحدار (3.3.I) يتيح لنا تقدير MPC وبالتالي تقدير M .

فبتقدير الـ MPC يمكن التنبؤ بالدخل ، الإنفاق الاستهلاكي ، العمالة ، . . . الخ . ويتبع ذلك التغير في السياسات الحكومية .

8- استخدام نموذج المراقبة : Use of the model for control or Policy Purposes

إذا اعتبرنا دالة الاستهلاك المقدرة في (3.3.I) ، وبافتراض أن الحكومة تعتقد أن الإنفاق الاستهلاكي حوالي 4900 (بليون دولار في سنة 1992) سوف يؤدي إلى معدل بطالة 4.2% (قبل سنة 2000) . وهنا يكون من الأهمية الإجابة عن السؤال التالي : ماهو مستوى الدخل الذي يضمن تحقيق الكمية المستهدفة للإنفاق الاستهلاكي 4900 بليون دولار؟

وباستخدام نموذج في (3.3.I) وبالتعويض عن $\hat{Y}= 4900$ نجد أن :

$$4900 = -184.0779 + 0.7064X \qquad (6.3.\text{I})$$

وبحل المعادلة (6.3.I) نجد أن $X= 7197$ بليون دولار تقريباً . وهذا يعني أن مستوى الدخل 7179 بليون دولار ، بشرط أن MPC تساوي 0.70 سوف يؤدي إلى إنفاق 4900 بليون دولار .

ومن ثم ، يمكن استخدام نموذج الانحدار للمراقبة أو متابعة السياسات . حيث يمثل المتغير المستقل X متغير مراقبة X control variable ويمثل المتغير Y المتغير المستهدف Y target variable .

وشكل (4) يلخص تحليل النمذجة القياسية التقليدية

Economic theory
↓
Mathematical model of theory
↓
Econometric model of theory
↓
Data
↓
Estimation of econometric model
↓
Hypothesis testing
↓
Forecasting or prediction
↓
Using the model for control or policy purposes

شكل (4) تحليل النمذجة القياسية

المقارنة بين النماذج المتنافسة : Choosing among competing models

عادة تقوم جهة حكومية بجمع البيانات الاقتصادية(على سبيل المثال ، قسم التجارة بالولايات المتحدة) مثل البيانات في جدول (1) . فعملية جمع البيانات لاتتطلب الأخذ في الاعتبار النظرية الاقتصادية ، ولكنها تقوم برصد الظاهرة في صورة كمية فقط .

ومما هو جدير بالذكر ، أنه يمكن بناء أكثر من نموذج لنفس البيانات . فعلى سبيل المثال ، طور Milton Friedman نموذج الاستهلاك بإضافة فرض الدخل الدائم permanent income hypothesis[15] Robert Hall نموذج الاستهلاك أيضاً بإضافة فرض دورة حياة الدخل الدائم life cycle permanent income hypothesis[16] .

(15)Milton Friedman, A Theory of Consumption Function, Princeton Univesity Press Princeton N.J. 1957.

(16)R. Hall, "Stochastic Implications of the Life Cycle Penmanent Income Hypothesis. Theory and Evidence," Journal of Political Economy 1978 vol 86 pp 971-987

والمشكلة التي تواجه الباحث علمياً هي كيف يمكن تحديد النموذج من النماذج المتنافسة لتمثيل الظاهرة محل الدراسة . ويرى Miller أن النموذج الأفضل هو النموذج الأكثر اتساقاً مع البيانات(17) .

كذلك يرى Clive Granger أنه عند استخدام النموذج في التنبؤ لابد من الأخذ في الاعتبار السؤالين التاليين :

1 - ما الهدف من النموذج؟- وماهي القرارات الاقتصادية المؤدية إلى تحقيق هذا الهدف؟

2 - هل توجد وسيلة يمكن باستخدامها إجراء مقارنة متكافئة بين النماذج المتنافسة؟

ويرى المؤلف أن هذه الأسئلة تعتبر ضرورية في البحث والمناقشة الاقتصادية . وخلال الأبواب المختلفة في هذا الكتاب ، سوف نعرض عدداً من النماذج المتنافسة كمحاول لتوضيح الظواهر الاقتصادية المختلفة .

فعلى سبيل المثال ، الطلاب المتخصصون في الاقتصاد ملمون بمفهوم دالة الإنتاج production function التي تعتبر أساساً علاقة بين المخرج output والمدخلات inputs (مثلاً رأس المال capital والعمالة labor) . وفي الأدب توجد دالتان للإنتاج معرفتان جيداً هما : دالة كوب دوجلاس Cobb- Douglas ، ودالة المرونة الثابتة constant elasticity . ووفقاً لبيانات المخرج ، وبيانات المدخلات يمكن استنتاج أي الدالتين تعتبر توفيق fitting أفضل .

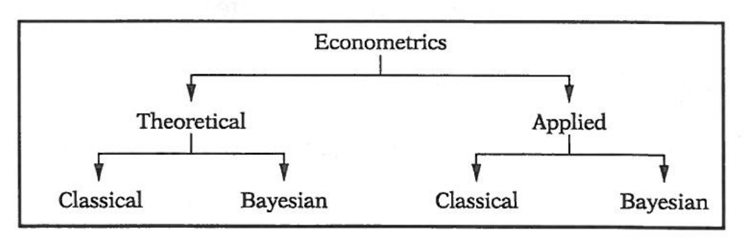
وهنا يجب طرح السؤال التالي : هل ممكن تطوير منهجية الاقتصاد القياسي (الثماني خطوات المتتالية السابق تناولها) بحيث تتضمن الفروض (النماذج) المتنافسة؟ وسوف نتناول ذلك في الباب (13) ، بعد تناول أساسيات نظرية الاقتصاد القياسي econometric theory .

---

(17)R. W. Miller, Fact and Method: Explanation, Confirmation, and Reality in the Natural and Soaal Scrences, Pnnceton University Press, Princeton, N.J., 1978, p. 176.

4- فروع الاقتصاد القياسي : TYPES OF ECONOMETRICS

ويمكن تقسيم الاقتصاد القياسي إلى فرعين أساسيين هما : الاقتصاد القياسي النظري theoretical econometrics ، والاقتصاد القياسي التطبيقي applied econometrics وكل فرع يمكن تقسيمه وفقاً للأساليب إلى فئتين هما : فئة الأساليب التقليدية classical techinque ، وفئة الأساليب البيزية Bayesian techinques كما هو موضح بشكل (5) .



شكل (5) يوضح ما هو الاقتصاد القياسي النظري والقياسي التطبيقي

ويهتم الاقتصاد القياسي النظري بالطرق المناسبة لقياس العلاقات الاقتصادية من خلال نماذج الاقتصاد القياسي و econometric models . وفي هذا الجانب ، يكون التركيز على الإحصاء الرياضي mathematical statistics . فعلى سبيل المثال ، تعتبر طريقة المربعات الصغرى least square إحدى الطرق المستخدمة في هذا الكتاب . فبالنسبة لهذه الطريقة مثلاً يتناول الاقتصاد القياسي النظري الفروض assumptions ، وخصائص الحل properties of solution ، وتأثير تغير أحد هذه الفروض أو أكثر على خصائص الحل . أما الاقتصاد القياسي التطبيقي ، فهو يستخدم أدوات tools الاقتصاد القياسي النظري في دراسة مجال خاص معين في الاقتصاد economics أو الأعمال business مثل دالة الإنتاج production function ، دالة الاستثمار investment function ، دوال العرض والطلب demand and supply functions ، نظرية حافظة الأوراق المالية portfolio theory ، . .الخ .

ويهتم هذا الكتاب بتنمية طرق الاقتصاد القياسي من خلال الفروض والاستخدامات والمحددات limitations لهذه الطرق . مع توضيح هذه الطرق من

خلال مجموعة متنوعة من الأمثلة في مناطق اقتصادية مختلفة . ورغم ذلك ، فإن هذا الكتاب لايقدم الاقتصاد القياسي التطبيقي بمعنى الدراسة الدقيقة المتعمقة في مجال التطبيق الاقتصادي . ولكن توجد مراجع متخصصة تتناول ذلك ، وتتضمن قائمة المراجع بهذا الكتاب مجموعة من هذه المراجع .

5 - المتطلبات الرياضية والإحصائية:
mathematical and statistical prerequisites

ورغم أن مستوى الكتاب أولي elementary level ، إلا أن الكاتب يفترض أن القارئ ملم بالمفاهيم الأساسية للتقدير الإحصائي statistical estimation ، واختبارات الفروض hypothesis testing . ويوجد بملحق A بالكتاب ، عرض شامل لأهم المفاهيم الإحصائية الأساسية ، كذلك في ملحق B يوجد ملخص لأهم أساسيات جبر المصفوفات matrix algebra ، كذلك ملحق C يتضمن ملخصًا لأساسيات نظرية الانحدار basic regression theory .

6 - دور الحاسب : THE ROLE OF THE COMPUTER

يعتبر تحليل الانحدار regression analysis من الأدوات المهمة والجيدة في الاقتصاد القياسي . واستخدام الحاسب يعتبر ضرورة في تحليل الانحدار أو التحليل الإحصائي بشكل عام . وتوجد حزم جاهزة للانحدار regression packages متاحة تجارياً مثل : Et: L imdep, SHAZAM, Micro TSP, Minitab, Eviews, SAS, SPSS, stata, Microbit, pcgive, BMD .

حيث تتناول هذه الحزم الأساليب المعروضة بهذا الكتاب .

في هذا الكتاب ، سوف يتدرب القارئ من حين لآخر على إجراء تجارب مونت كارلو(*) Monte Carlo experiments بحزمة إحصائية أو أكثر – وذلك بهدف فهم

(*) مونت كارلو – أحد أساليب المحاكاة Simulation technique .

خصائص الطرق الإحصائية المقدمة في الكتاب . وسوف نناقش تفاصيل تجارب مونت كارلو في موضعها في الكتاب .

7 - مقترحات لقراءات إضافية :
SUGGESTION FOR FURTHER READING

موضوع منهجية الاقتصاد القياسي موضوع ضخم وقابل للمناقشة . ولأهمية هذا الموضوع ، أقترح الكتب التالية لمزيد من الاطلاع في هذا الموضوع :

- في سنة 1989 قدم Neil de Marchi and Christopher كتابًا تحت عنوان تاريخ ومنهجية الاقتصاد القياسي History and Methodology of Econometrics .

- في سنة 1997 قدم Wojciech W.charemza and Derek F.Deadman كتابًا تحت عنوان "الاتجاهات الحديثة في الاقتصاد القياسي New Directions in Econometric" .

كذلك مؤلفو هذا الكتاب ، قدموا أساليب حديثة في منهجية الاقتصاد القياسي :

- في سنة 1990 قدم Adriam C.Darnell and J.lynne Evans كتابًا تحت عنوان "محددات الاقتصاد القياسي The limits of econometrics" .

- في سنة 1990 قدم Mary S.Morgan كتابًا تحت عنوان "تاريخ فكر الاقتصاد القياسي The history of econometric ideas" .

- في سنة 1995 قدم David F.Hendry and Mary S.Morgan كتابًا تحت عنوان "أسس تحليل الاقتصاد القياسي The foundation of econometric analysis"

وتعتبر الكتب التالية ذات فائدة مهمة في موضوعات الإحصاءات البيزية Bayesian statistics والاقتصاد القياسي .

- في سنة 1985 قدم John H.dey كتابًا تحت عنوان "بيانات في اشتباه Data in Doubt" .

• في سنة 1989 قدم Peter M.lee كتابًا تحت عنوان "الإحصاءات البيزية Bayesian statistics" .

• في سنة 1995 قدم Dale J.Brier كتابًا تحت عنوان "المستوى المتوسط للإحصاء والاقتصاد القياسي Intermediate statistics and Econometrics" .

• في سنة 1971 قدم Arnold Zeller كتابًا تحت عنوان "مقدمة في الاستدلال البيزي في الاقتصاد القياسي An introduction to Bayesian inference in Econometrics" .

مقدمة المترجم

يعتبر هذا الكتاب من الكتب المهمة والأساسية في الاقتصاد القياسي . ويعتبر الاقتصاد القياسي في الأساس توليفة من النظرية الاقتصادية والإحصاء حيث إن أساليب البحث والتحليل في الاقتصاد القياسي تعتمد أساسًا على استخدام الاستدلال الإحصائي للربط بين النظرية الاقتصادية وفروضها من جهة وقياسات القيم الفعلية التي يتم الحصول عليها في الواقع العملي من الجهة الأخرى .

هذا الكتاب ينقسم إلى جزءين : الجزء الأول ويشتمل على ثلاثة عشر فصل يتم فيها تناول كل الجوانب المتعلقة بنماذج الانحدار الخطية بداية من النماذج البسيطة التي تحتوى على متغير مفسر واحد ثم متغيرين أثنين ثم ثلاث متغيرات حتى نصل إلى نماذج الانحدار المتعدد كما يتم استعراض كافة فروض نموذج الانحدار التقليدية إلى جانب كافة التفاصيل الخاصة بالمشاكل العملية التي تظهر عند عدم تحقق مثل هذه الفروض . يتميز الكتاب بالعديد من الأمثلة التطبيقية المطولة ذات البيانات الحقيقية حيث يتم تحليل الجانب الاقتصادي فيها بشكل إحصائي دقيق بالإضافة إلى استعراض العديد من الحزم الإحصائية التي تستخدم الأساليب المعروضة بهذا الكتاب .

أما الجزء الثاني من الكتاب يشتمل على تسعة فصول يتم فيها تناول نماذج الانحدار غير الخطية إلى جانب النماذج ذات المتغيرات الوهمية ومن المواضيع الحيوية التي يتم تناولها في الجزء الثاني نماذج الاقتصاد الديناميكية والمرتبطة ارتباط وثيق بتحليل السلاسل الزمنية وما تحتويه من اختبارات وتحويلات وأساليب تنبؤ .

هذا الكتاب يعتبر كتاب مرجعي في الاقتصاد القياسي ويتطلب حد أدنى من الخلفية الإحصائية والإلمام بمفاهيم التقدير الإحصائي واختبارات الفروض ويتم عرض المفاهيم الإحصائية الأساسية بالإضافة إلى جبر المصفوفات وبعض الخلفيات الرياضية للانحدار في ملاحق هذا الكتاب .

لأهمية هذا الكتاب طلب مني ترجمة ومراجعة هذا الكتاب باللغة العربية وأود أن أتقدم بوافر الشكر إلى الأستاذة الدكتورة عفاف الدش ، أستاذة الاحصاء بكلية التجارة – جامعة حلوان على مساهمتها بترجمة الكتاب من الصفحة 49 إلى الصفحة 210 .

وأتمنى أن أكون قد وفقت في ترجمة الأفكار والمفاهيم المعقدة الموجودة في الكتاب إلى اللغة العربية بدرجة وافية من الوضوح والدقة .

أ.م.د. هند عبدالغفار عودة

رئيس قسم الرياضة والإحصاء

كلية التجارة – جامعة حلوان

الفصل الأول

نماذج انحدار معادلة المتغير الواحد(*)

SINGLE- EQUATION REGRESSION MODELS(*)

الفصل الأول من هذا الكتاب، يتناول نماذج الانحدار في حالة معادلة المتغير الواحد. في هذا النوع من النماذج يوجد متغير واحد تابع يعبر عنه في شكل دالة خطية في متغير واحد أو أكثر من المتغيرات المستقلة . في هذه النماذج يفترض وجود العلاقات السببية .

الفصل الأول: نناقش - وفقاً للتطور التاريخي - التفسير الحديث لتعبير انحدار regression، ونوضح الفرق بين تفسيرين، مع تقديم أمثلة اقتصادية وغير اقتصادية متنوعة .

الفصل الثاني: يقدم المفاهيم الأساسية لتحليل الانحدار من خلال نموذج انحدار المتغيرين، متغير تابع واحد، ومتغير مستقل (مفسر) واحد .

الفصل الثالث: وهو امتداد نموذج المتغيرين، أحدهما تابع، والآخر مستقل (مفسر)، والمعروف بنموذج الانحدار الخطي التقليدي حيث يبنى هذا النموذج على مجموعة من الافتراضات البسيطة . وفي ضوء هذه الفروض، سوف تقدم طريقة المربعات الصغرى العادية ordinary least squares (OLS) كتقدير معلمات نموذج المتغيرين .

(*) الكتاب من ص 49 إلى ص 210 ترجمة أ .د .عفاف الدش .

طريقة المربعات الصغرى العادية (OLS)، واستخدامها يرجع إلى الحصول على تقديرات معلمات النموذج لها خصائص إحصائية مرغوب فيها.

الفصل الرابع: يتناول نموذج الانحدار في متغيرين، يسمى بنموذج الانحدار الطبيعي التقليدي، ويبنى هذا النموذج على افتراض أن المتغير التابع متغير عشوائي يتبع التوزيع الاحتمالي الطبيعي. وتحت هذا الفرض، نجد أن استخدام طريقة المربعات الصغرى العادية (OLS) تعطي تقديرات لها خصائص جيدة، ولايتحقق ذلك في حالة عندما يكون المتغير التابع متغيراً لايتبع التوزيع الطبيعي. وأهمية هذه الخصائص ترجع إلى استخدامها في اختبارات الفروض.

الفصل الخامس: ويتناول اختبارات الفروض. وذلك بمعنى هل التقديرات التي تم الحصول عليها متلائمة مع القيم الافتراضية المقترحة باستخدام النظرية أو عن طريق عمل تكراري سابق.

الفصل السادس: ويعتبر امتداداً لنموذج المتغيرين، وبصفة خاصة في الموضوعات التالية:

1 - الانحدار خلال الأصل.

2 - التدرج ووحدات القياس:

3 - الصيغ الدالية لنماذج الانحدار مثل اللوغاريتم المزدوج، شبه اللوغاريتم، النماذج العكسية التبادلية.

الفصل السابع: ويتناول نماذج الانحدار المتغير، حيث يوجد متغير واحد تابع وأكثر من متغير واحد مستقل (أو مفسر). ونوضح كيف يمكن استخدام طريقة المربعات الصغرى العادية (OLS) للحصول على تقديرات المعلمات للنموذج.

الفصل الثامن: وهو يعتبر امتداداً للفصل الخامس، حيث يتناول المفاهيم في الفصل الخامس بالنسبة لنموذج الانحدار المتعدد. حيث يتناول بعض الصعوبات الناشئة عن تناول أكثر من متغير واحد مستقل (مفسر).

الفصل التاسع: ويتناول النماذج التي تتضمن متغيرات وصفية.

فهذا الفصل يؤكد على أنه ليس بالضرورة أن تكون جميع المتغيرات المستقلة (المفسرة) كمية، ولكن يمكن أن يكون بعضها وصفيًا (نوعيًا) مثل النوع، العرق، الديانة ، الجنسية، منطقة الإقامة . فهذه المتغيرات الوصفية (النوعية) تلعب دوراً مهمًا في تفسير كثير من الظواهر الاقتصادية .

طبيعة تحليل الانحدار :
THE NATURE OF REGRESSION ANALYSIS

كما ذكرنا في المقدمة، أن الانحدار regression أداة رئيسة في الاقتصاد القياسي . وفي هذا الفصل سوف نقدم باختصار طبيعة هذه الأداة .

1.1 الأصل التاريخي لمصطلح انحدار :
HISTORICAL ORIGIN OF THE TERM REGRESSION

قدم Francis Galton سنة 1886 اصطلاح انحدار في مقاله المشهور الذي وضح فيه أنه بالرغم من وجود استعداد (ميل) للآباء الطوال إنجاب أبناء طوال، كذلك الآباء القصار إنجاب أطفال قصار أيضاً، فإن متوسط الطول للأطفال من آباء بأطوال معينة تنحدر تجاه متوسط الطول في السكان ككل[1] . أو بعبارة أخرى، طول الأطفال لآباء بأطوال غير عادية (طول أو قصر) تميل إلى متوسط الطول في المجتمع .

قانون Galton للانحدار العام تأكد بواسطة صديقه سنة 1903 . عندما جمع بيانات عن 1000 فرد من عائلات[2] مصنفة في فئات للأطوال . ووجد أن متوسط طول الأبناء لآباء أطوالهم في الفئات الأكثر طولاً كان أقل من طول الآباء، كذلك متوسط طول الأبناء لآباء أطوالهم في الفئات الأقل طولاً (القصار) كان أكبر من طول الآباء، هكذا صفة انحدار الأبناء الطوال أو القصار تتعادل مع متوسط الطول لجميع الأفراد .

في عبارات Galton كانت عبارة "الانحدار للوسطية " .

---

(1) Francis Galton, "Family Likeness in Stature," Proceedings of Royal Society, London, vol. 40, 1886. pp. 42–72.

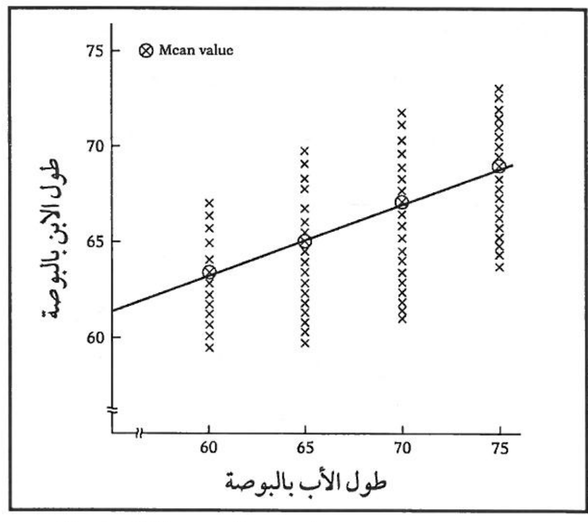
(2) K. Pearson and A. Lee, "On the Laws of Inheritance," Biometrika, vol. 2, Nov. 1903, pp. 357–462.

2.1 التفسير الحديث للانحدار:
THE MODERN INTERPRETATION OF REGRESSION

يهتم تحليل الانحدار بالتبعية لمتغير ويسمى بالمتغير التابع لمتغير أو أكثر يسمى بالمتغير (أو المتغيرات) المستقلة (أو المتغيرات المفسرة). ويستخدم الانحدار في التقدير والتنبؤ لمتوسط قيم المتغير التابع عند قيم معينة للمتغيرات المستقلة. وخلال عرض تحليل الانحدار، سوف تتضح أهميته. وفيما يلي سوف نقدم بعض الأمثلة البسيطة لتوضيح المفاهيم الأساسية.

أمثلة : Examples

1 – بإعادة النظر في قانون Golton للانحدار، نجد أن Galton اهتم باستنتاج لماذا يوجد استقرار في توزيع الأطوال في المجتمع. ولكن وجهة النظر الحديثة لم تهتم بالعرض ولكن بكيفية استنتاج أن متوسط طول الأبناء يتغير عند طول معين للآباء. وشكل الانتشار في شكل (1.1) يوضح ذلك.

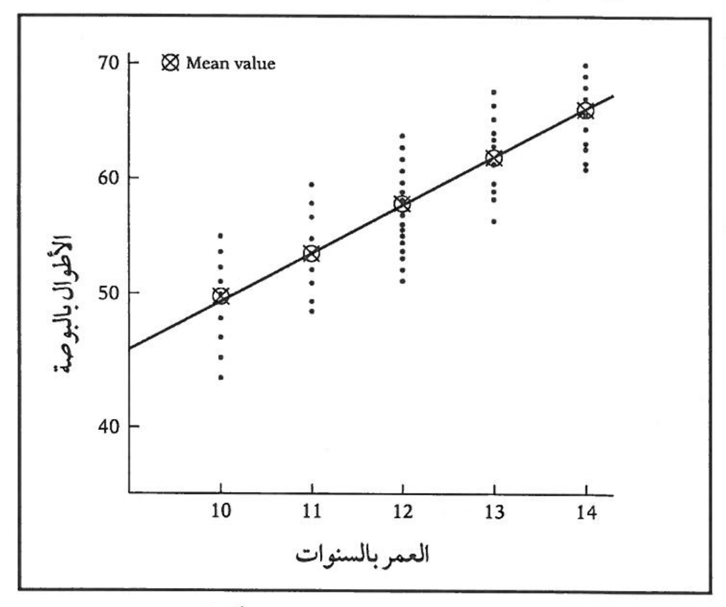


شكل (1.1) توزيع افتراضي لأطوال الأبناء وفقاً لأطوال الآباء

ويوضح الشكل توزيع أطوال الأبناء في مجتمع افتراضي وفقاً لقيم محددة لأطوال الآباء. ونلاحظ أنه وفقاً لأي طول معين للأب يوجد توزيع لأطوال الأبناء. كذلك لوحظ أنه بالرغم من أن أطوال الأبناء عند قيمة معينة لطول

الأب، فان الوسط الحسابي (المتوسط) لطول الأبناء بصفة عامة يتزايد كلما تزايد طول الأب. وفي الشكل السابق، نجد أن العلامة (+) تشير إلى متوسط طول الأبناء عند طول معين للأب. وبتوصيل هذه النقاط (+) التي تشير إلى متوسطات أطوال الأبناء عند أطوال معينة للآباء نحصل على خط الانحدار كما هو موضح بالشكل (3.1) [3].

2 - اعتبر شكل الانتشار التالي، حيث يوضح الشكل توزيع أطوال (بالبوصة) مجموعة من الأولاد في مجتمع افتراضي وفقاً للعمر (بالسنوات). فوفقاً لعمر معين يوجد توزيع للأطوال.

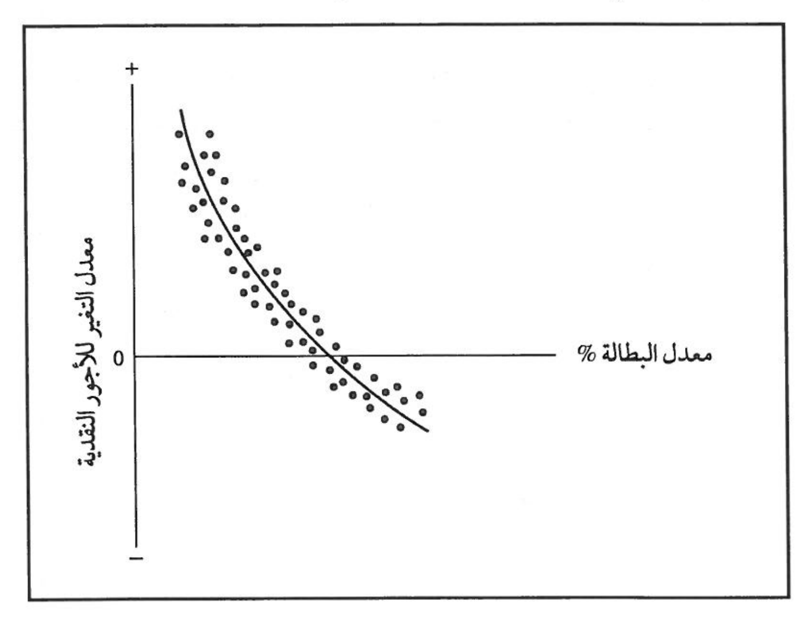


شكل (2.1) توزيع افتراضي للأطوال وفقاً للأعمار

ويلاحظ أنه ليس كل الأولاد عند عمر معين متماثلين في الطول. ولكن متوسط الطول يتزايد بزيادة العمر (بالطبع إلى عمر معين). وخط الانحدار في الشكل يمثل متوسط الطول عند عمر معين. هكذا بمعرفة العمر يمكن التنبؤ بمتوسط الطول باستخدام خط الانحدار.

(3) At this stage of the development of the subject matter, we shall call this regression line simply the line connecting the rnean, or average, value of the dependent variable (son's height) corresponding to the given val~e of the explanatory variable (father 3 height). Note that this line has a positive slope but the slope is less tham 1, which is in conformity with Galton's regression to mediocrity (why?)

3 - ومن الأمثلة الاقتصادية - اهتمام الاقتصادي بتتابعية الإنفاق الاستهلاكي على الدخل الشخصي الحقيقي. وتحليل الانحدار يكون مفيداً في تقدير الميل الحدي للاستهلاك. أو بعبارة أخرى، متوسط تغير الإنفاق الاستهلاكي الناتج عن تغير الدخل الحقيقي بدولار واحد (انظر شكل 3.1).

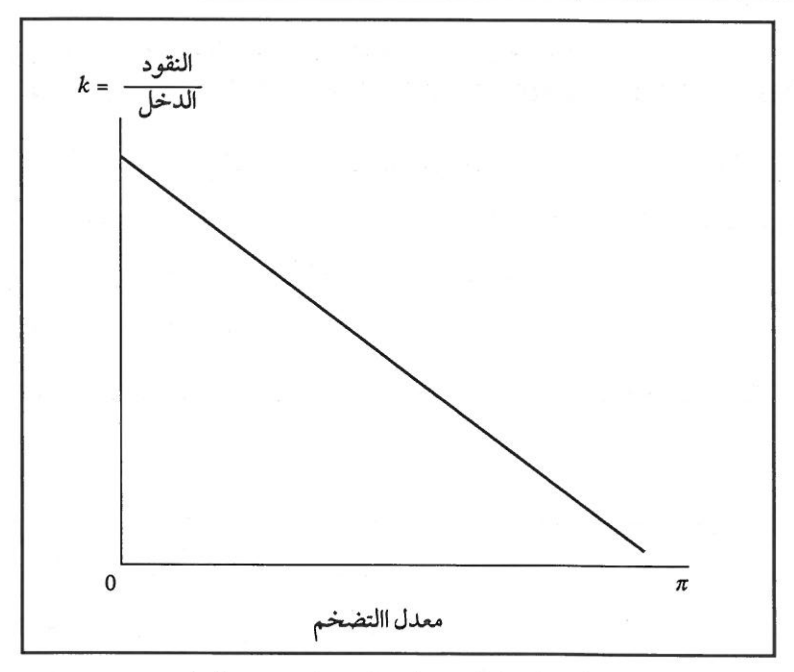


شكل (3.1) متوسط تغير الإنفاق الاستهلاكي

4 - وبالنسبة للسوق الاحتكارية، يكون من الأهمية بالنسبة للمحتكر أن يقوم بتحديد السعر أو الكمية المعروضة من المنتج (وليس الاثنين معاً) عادة يرغب في تقدير المرونة السعرية للطلب بهدف تحديد السعر الأكثر ربحية.

5 - في الاقتصاد الحر - عادة يرغب الاقتصادي في دراسة معدل التغير للأجور النقدية وعلاقتها بمعدل البطالة. وشكل الانتشار في شكل (3.1) يوضح معدل تغير الأجور النقدية بالنسبة لمعدل البطالة، وهذه المعرفة مفيدة في تحديد بداية ما لعملية التضخم في الاقتصاد.

فزيادة الأجور النقدية تنعكس في زيادة الأسعار.



شكل (4.1) النقود المتداولة وعلاقتها بمعدل التضخم

6 - من الاقتصاد النقدي يتضح أنه عند ثبات باقي العوامل، فإن أعلى معدل للتضخم يكون عند أقل نسبة k ، حيث k عبارة عن كمية النقود إلى الدخل كما في شكل (4.1). والتحليل الكمي لهذه العلاقة يمكن الاقتصادي من التنبؤ بكمية النقود كنسبة إلى الدخل.

7 - في الشركات الإنتاجية عادة يرغم مدير التسويق في تحديد العلاقة بين الكمية المطلوبة من المنتج، وحجم المخصص في الميزانية للإعلان عن المنتج.

وبإيجاد مرونة الطلب بالنسبة للمخصص على الإنفاق على الإعلان يمكن تحديد القيمة المثلى المخصصة للإعلان عن المنتج في الميزانية.

كذلك يهتم المهندسون الزراعيون بدراسة العلاقة بين المحصول مثل القمح مثلاً وكل من درجة الحرارة، ومعدل سقوط الأمطار، كمية السماد، الخ فتحليل التابعية يمكن من التنبؤ بمتوسط كمية المحصول عند معلومات معينة عن المتغيرات المفسرة.

3.1 العلاقات الإحصائية مقابل العلاقات اليقينية:
STATISTICAL VERSUS DETERMINISTIC RELATIONSHIPS

من الأمثلة السابقة في الفصل (2.1) يلاحظ القارئ أن تحليل الانحدار يهتم بعلاقة المتغير التابع بالمتغيرات المفسرة. حيث تكون العلاقة علاقة إحصائية [4] وليست علاقة يقينية. فالعلاقات الإحصائية تعتمد على متغيرات عشوائية، أي متغيرات لها توزيعات احتمالية.

وفيما يلي، سوف نوضح الفرق بين العلاقة الإحصائية والعلاقة اليقينية على النحو التالي:

إنتاجية الفدان من محصول معين تعتمد على درجة الحرارة، كمية المياه، كمية السماد، . . الخ. فنجد أن إنتاجية الفدان من محصول ما تعتبر متغيراً عشوائيًا، حيث لايستطيع المزارع التنبؤ برقم معين للإنتاجية عند قيم محددة لكل متغير من المتغيرات المفسرة (المياه، الحرارة، السماء، . . الخ). حيث عند قيم محددت للمتغيرات المفسرة توجد أكثر من قيمة للمتغير التابع (إنتاجية الفدان) وذلك يرجع إلى وجود مجموعة من العناصر أو العوامل لها تأثير تجميعي (أي تأثيرها معاً) على المتغير التابع، ويصعب قياسها بل قد لايمكن قياسها. وتسمى هذه العناصر بالعوامل العشوائية أو الأخطاء. وفقاً لذلك، يكون المتغير التابع دالة في المتغيرات المفسرة والمتغيرات العشوائية، وتسمى العلاقة في هذه الحالة علاقة إحصائية.

أما العلاقات اليقينية فهي مثلاً مثل قانون نيوتن Newton للجاذبية، حيث نجد أن قوة الجاذبية بين جسمين ونرمز لها بالرمز F تعتمد على كتلة كل جسم ولتكن m_1 و m_2 والمسافة بين الجسمين ولتكن r وتأخذ الشكل التالي:

$$F = k(m_1 m_2 / r^2)$$

حيث k مقدار ثابت، فنجد أن قيمة محددة لكل من k ، $r, m_1 m_2$ توجد قيمة محددة لـ F. أي لاتوجد متغيرات عشوائية تؤثر على المتغير التابع F، وبالتالي تسمى

(4) The word stochastic comes from the Greek word stokhos meaning "a bull's eys," The outcome of throwing darts on a dart board is a stochastic process, that is, a process fraught wit misses.

العلاقة في هذه الحالة بالعلاقة اليقينية. كذلك من العلاقات اليقينية قانون الأول ohm's law ، أي قانون وحدة المقاومة الكهربائية، حيث يتم تحديد مقاومة الموصلات المعدنية metallic conductors على النحو التالي :

$$C = \frac{1}{k} V$$

حيث V تشير إلى الفولتات، $\frac{1}{k}$ مقدار ثابت. فنجد أن العلاقة بين المتغير التابع C والمتغير المستقل V علاقة يقينية.

4.1 الانحدار مقابل السببية : REGRESSION VERSUS CAUSATION

رغم أن تحليل الانحدار يتعامل مع علاقة المتغير التابع بالمتغيرات الأخرى المفسرة، ولكنه لايؤدي إلى السببية. فكما ذكر كل من Kendall and Stuart " العلاقة الإحصائية على الرغم من قوتها ودلالتها، إلا أنها لا تؤكد الاتصال السببي بين المتغير التابع والمتغيرات المفسرة. وتفكيرنا في السببية يجب أن يأتي من خارج نطاق الإحصاء، وفي النهاية يكون من نظرية ما أو من أشياء أخرى[5] ".

وسوف نوضح ذلك من خلال الأمثلة السابقة، فمثلاً في مثال إنتاجية الفدان، نجد أن افتراض أن الإنتاجية ترجع إلى سقوط الأمطار كأحد العوامل، ويرجع ذلك إلى اعتبارات أخرى غير إحصائية، كذلك عدم اعتماد سقوط الأمطار على الإنتاجية غير مفسر إحصائياً. وبالتالي فإن العلاقة الإحصائية في حد ذاتها لاتؤدي إلى أسباب منطقية أو لا تؤدي إلى السببية[6].

5.1 الانحدار مقابل الارتباط: REGRESSION VERSUS CORRELATION

ودائماً يقترن تحليل الانحدار بتحليل الارتباط رغم الاختلاف بينهما. فتحليل الارتباط يهدف أساساً إلى قياس شدة (أو درجة) العلاقة الخطية بين المتغيرات. وسوف نتناول مقياس معامل الارتباط في الفصل الثالث. ومعامل الارتباط مقياس

(5) M. G. Kendall and A. Stuart, The Advanced Theory of Stafistics, Charles Griffn Publishers, NewYork, 1961,vol. 2, chap. 26, p. 279.

(6) But as we shall see in Chap. 3, classical regression analysis is based on the assumption that the model used in the analysis is the correct moadel. Therefore, the direction of causality may be implicit in the model postulated.

يقيس شدة العلاقة. فعلى سبيل المثال، قوة العلاقة بين التدخين وسرطان الرئة، كذلك درجات الطلاب في الإحصاء وارتباطها بدرجاتهم في الرياضيات[7].

ويوجد اختلاف أساسي بين الانحدار والارتباط، فتحليل الانحدار يهتم بالتعبير عن المتغير التابع كدالة في المتغيرات المفسرة. أما تحليل الارتباط فيهتم بقوة العلاقة بين المتغير التابع والمتغير المستقل. كذلك في تحليل الانحدار يكون المتغير التابع متغيراً عشوائيًا والمتغير المفسر (أو المتغيرات المفسرة) متغيرات غير عشوائية (يقينية)، في حين أنه يمكن في تحليل الارتباط أن يكون المتغيران عشوائيين[8].

6.1 المنهجية والرموز: TERMINOLOGY AND NOTATION

قبل تناول تحليل الانحدار، سوف نقدم في الجدول التالي بعض الاصطلاحات الواردة في الأدبيات لكل من المتغير التابع، والمتغير المفسر في تحليل الانحدار.

Dependent variable	Explanatory variable
⇕	⇕
Explained variable	Independent variable
⇕	⇕
Predictand	Predictor
⇕	⇕
Regressand	**Regressor**
⇕	⇕
Response	Stimulus
⇕	⇕
Endogenous	Exogenous
⇕	⇕
Outcome	Covariate
⇕	⇕
Controlled variable	Control variable

وفي هذا الكتاب، سوف نستخدم المتغير التابع، والمتغير المفسر. ودراسة التبعية لمتغير تابع، وآخر متغير واحد مفسر مثل الإنفاق الاستهلاكي (المتغير التابع)، والدخل (المتغير المفسر) يسمى تحليل الانحدار بتحليل انحدار متغيرين، أو تحليل

(7) It is crucial to note that the explanatory variables may be intrinsically stochashc but for the purpose of regression analysis we assume that their va ues are fixed in repeated sampling (that is, X assumes the same values in various samples), thus rendering them in effect non random or nonstochastic. But more on this in Chap. 3, Sec. 3.2.

(8) In advanced treatment of econometrics, one can relax the assumption that the explanatory variables are nonstochastic (see introduction to Part II).

الانحدار البسيط . وفي حالة وجود متغير تابع واحد وعدة متغيرات مفسرة يسمى تحليل الانحدار في هذه الحالة بتحليل الانحدار المتعدد .

وعادة يرمز للمتغير التابع بالرمز Y، كذلك المتغيرات X، حيث X متجه من المتغيرات التي عددها K $(X_1, X_2, ..., X_k)$ حيث يشير X_k إلى المتغير المفسر رقم k. كذلك إذا كان عدد المشاهدات في العينة يساوي t، والرمز يشير إلى المشاهدة رقم i، t ، $i = 1, 2, ...,$ ، فإن X_{ki} يشير إلى المشاهدة رقم i للمتغير X_k. وبصفة عامة، فإن X_{ji} تشير إلى المشاهدة رقم i للمتغير رقم j حيث $j = 1, 2, .., k$. والدليل رقم (i) تستخدم كبيان قطع عرضي (بمعنى بيانات مجمعة في نقطة زمنية) كذلك t تشير إلى بيانات السلسلة الزمنية (بمعنى بيانات مجمعة فوق فترة زمنية معينة). وطبيعة البيان العرض وبيانات السلسلة الزمنية يعتبر موضوعًا ذا أهمية بالنسبة لطبيعة ومصادر البيانات بالنسبة للتحليل التكراري، وسوف نناقش ذلك في الفصل التالي[9] .

7-1 طبيعة ومصادر البيانات في التحليل الاقتصادي:
THE NATURAL AND SOURCES OF DATA FOR ECONOMIC ANALYSIS[10]

يتوقف نجاح التحليل الاقتصادي على ملاءمة وإتاحة البيانات اللازمة. ولذلك من الأهمية تناول طبيعة ومصادر البيانات .

أنواع البيانات : Types of Data

يمكن تقسيم البيانات إلى ثلاثة أنواع هي :

- بيانات السلاسل الزمنية
- بيانات القطع العرضي
- البيانات المزدوجة (وتعني توليفة من بيانات السلاسل الزمنية ، وبيانات القطع العرضي) .

بيانات السلاسل الزمنية يعطي جدول (1) في المقدمة، مثالاً لبيانات السلسلة الزمنية. فالسلسلة الزمنية هي مجموعة من قيم المشاهدات لمتغير معين مأخوذة في فترات زمنية متتالية مثلاً يومية (مثل أسعار الأسهم، تقرير الطقس) أسبوعية (مثل

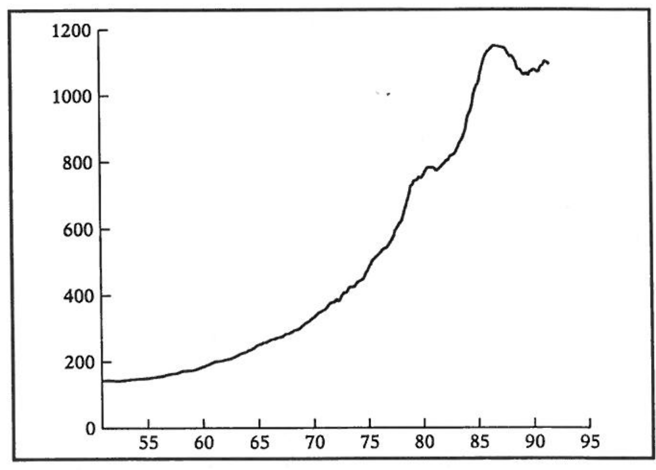
(9) See App. A for foumal definition and further details.

(10) For an informative account, see Michael D. Intriligator, Econometric Models, Techniques, and Applications, Prentice Hall, Englewood Cliffs, N. J., 1978, chap. 3

أسعار العملات)، شهرية (مثل معدلات البطالة)، ربع سنوية (مثل GDP)، سنوية مثل (الميزانيات الحكومية)، (خماسية) أي كل خمس سنوات (مثل المسوح في الصناعات). أو (عشرية) كل عشر سنوات (مثل المسوح السكانية).

وأحياناً تأخذ بيانات GDP أو الإنفاق الاستهلاكي كل ربع سنوية، ونظراً للتقدم الكبير في الحاسبات، فإنه يمكن تجميع البيانات في فترات زمنية صغيرة مثل بيانات أسعار الأسهم.

وتستخدم بيانات السلاسل الزمنية بكثافة في دراسات الاقتصاد القياسي، كما سوف نوضح في الفصول التالية المتعلقة بالسلاسل الزمنية في الاقتصاد القياسي. ولكن توجد مشكلة تواجه المتخصصين في الاقتصاد القياسي عند التعامل مع السلاسل الزمنية، وتتمثل هذه المشكلة في أن معظم الأعمال التكرارية مبنية على افتراض سكون السلسلة الزمنية، أي أن السلسلة الزمنية ساكنة والسكون له مدلول سوف نتناوله بالتفصيل فيما بعد، ولكن باختصار شديد، يمكن القول إن السلسلة الزمنية ساكنة إذا كان المتوسط والتباين للمتغير محل الدراسة لاتتغير بشكل منتظم خلال الزمن، أو بعبارة أخرى السلوك الإحصائي للمتغير لايتغير. ولتوضيح هذا المصطلح، إذا اعتبرنا شكل (5.1) حيث يوضح سلسلة زمنية للمعروض من النقود M1 في الولايات المتحدة اعتباراً من واحد يناير 1959 حتى واحد وثلاثين يوليو 1999 (البيانات الفعلية موجودة في تمرين 4.1).



شكل (5.1) عرض النقود M1 في الولايات المتحدة في الفترة 01 : 1951–9 : 1999

ويتضح من الشكل أن اتجاه trend المعروض من النقود يتغير خلال الزمن، وبالتالي، فإن السلسلة الزمنية في هذه الحالة تعتبر غير ساكنة[11]، وسوف نتناول هذا الموضوع بالتفصيل في الفصل (21) .

بيانات القطع العرضي : cross- section data

وبيانات القطع العرضي هي بيانات لمتغير أو أكثر جمعت في نفس النقطة الزمنية. ومثال ذلك، مسح السكان (تعداد السكان) عن طريق مكتب المسوح كل 10 سنوات (آخر هذه المسوح سنة 2000)، أومسوح الإنفاق الاستهلاكي التي أجرتها جامعة Michigan .

ومن الأمثلة الواضحة لبيانات القطع العرضي البيانات الموجودة بجدول (1.1) .

U.S. EGG PRODUCTION

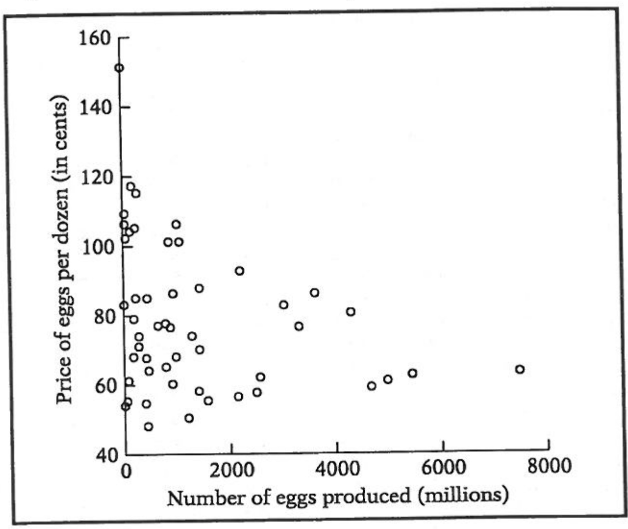
State	Y_1	Y_2	X_1	X_2	State	Y_1	Y_2	X_1	X_2
AL	2,206	2,186	92.7	91.4	MT	172	164	68.0	66.0
AK	0.7	0.7	151.0	149.0	NE	1,202	1,400	50.3	48.9
AZ	73	74	61.0	56.0	NV	2.2	1.8	53.9	52.7
AR	3,620	3,737	86.3	91.8	NH	43	49	109.0	104.0
CA	7,472	7,444	63.4	58.4	NJ	442	491	85.0	83.0
CO	788	873	77.8	73.0	NM	283	302	74.0	70.0
CT	1,029	948	106.0	104.0	NY	975	987	68.1	64.0
DE	168	164	117.0	113.0	NC	3,033	3,045	82.8	78.7
FL	2,586	2,537	62.0	57.2	ND	51	45	55.2	48.0
GA	4,302	4,301	80.6	80.8	OH	4,667	4,637	59.1	54.7
HI	227.5	224.5	85.0	85.5	OK	869	830	101.0	100.0
ID	187	203	79.1	72.9	OR	652	686	77.0	74.6
IL	793	809	65.0	70.5	PA	4,976	5,130	61.0	52.0
IN	5,445	5,290	62.7	60.1	RI	53	50	102.0	99.0
IA	2,151	2,247	56.5	53.0	SC	1,422	1,420	70.1	65.9
KS	404	389	54.5	47.8	SD	435	602	48.0	45.8
KY	412	483	67.7	73.5	TN	277	279	71.0	80.7
LA	273	254	115.0	115.0	TX	3,317	3,356	76.7	72.6
ME	1,069	1,070	101.0	97.0	UT	456	486	64.0	59.0
MD	885	898	76.6	75.4	VT	31	30	106.0	102.0
MA	235	237	105.0	102.0	VA	943	988	86.3	81.2
MI	1,406	1,396	58.0	53.8	WA	1,287	1,313	74.1	71.5
MN	2,499	2,697	57.7	54.0	WV	136	174	104.0	109.0
MS	1,434	1,468	87.8	86.7	WI	910	873	60.1	54.0
MO	1,580	1,622	55.4	51.5	WY	1.7	1.7	83.0	83.0

Note: Y_1 = eggs produced in 1990 (millions)
Y_2 = eggs produced in 1991 (millions)
X_1 = price per dozen (cents) in 1990
X_2 = price per dozen (cents) in 1991
Source: World Almanac, 1993, p. 119. The data are from the Economic Research Service, U.S. Department of Agriculture.

(11) To see this more clearly, we divided the data into four time periods: 1951:01 to 1962:12; 1963:01 to 1974:12,1975:01 to 1986:12 and 1987:01 to 1999:09: For these subperiods themean values of the money supply (with corresponding standard deviations in parentheses) were, respectively, 165.88 (23.27), 323.20 (72.66), 788.12 (195A:~), and 1099 (27.84), all figures in billions of dollars. This is a rough indication of the fact that the money supply over the entire period was not stationary.

فالجدول يوضح كميات البيض المنتجة (بالمليون بيضة) في 50 ولاية في عامي 1990، 1991، حيث Y_1، Y_2 تشيران إلى الكمية في 1990، 1991 على الترتيب، كذلك يتضمن الجدول سعر دستة (12 بيضة) البيض بالسنت في كل ولاية المناظرة للكميات المنتجة، حيث X_1، X_2 تشيران إلى السعر في 1990، 1991 على الترتيب. ففي سنة 1990 نجد لدينا متغيرين Y_1، X_1 تم قياسهما في نفس النقطة الزمنية.. بالمثل في سنة 1991 يوجد لدينا Y_1، X_1 تم قياسهما في نفس النقطة الزمنية. وتعاني بيانات القطع العرضي من نفس مشاكل السلسلة الزمنية وبصفة خاصة مشكلة التغاير heterogeneity. ومن الجدول نجد بعض الولايات تمتاز بكميات الإنتاج الكبيرة مثل (ولاية Pennsylvania) وولايات أخرى تمتاز بكميات الإنتاج القليلة (مثل ولاية Alasca).

وشكل (6.1) يوضح الاختلاف في الكميات المنتجة والأسعار في 50 ولاية



شكل (6.1) يوضح العلاقة بين الكميات المنتجة للبيض والأسعار في سنة 1990 في 50 ولاية

وشكل الانتشار في شكل (6.1) يوضح العلاقة بين الكميات المنتجة والأسعار سنة 1990. وفي الفصل (11) سوف نوضح كيفية تحديد العلاقة بين المتغيرات الاقتصادية economic variables.

البيانات المزدوجة: pooled data

البيانات المزدوجة هي مزيج من بيانات السلسلة الزمنية، وبيانات القطاع العرضي. وجدول (1.1) يعتبر مثالاً جيداً للبيانات المزدوجة أيضاً.

ففي كل سنة، نجد لدينا قطاعًا عرضيًا من البيانات متمثلاً في 50 مشاهدة تمثل الكمية والسعر في ولاية معينة. وكل ولاية يوجد عدد 2 سلسلة زمنية تمثل الكمية، وسلسلة تمثل السعر. وبالتالي يصبح لدينا 100 مشاهدة مركبة في العامين 1990، 1991 . وأيضاً بيانات تمرين (1.1) توضح البيانات المزدوجة (المركبة) للرقم القياسي لأسعار الاستهلاك (CPI) للدولة خلال الفترة 1973-1997 . كذلك توضح (CPI) لعدد 7 دول خلال نفس الفترة. وبالتالي في السنة الواحدة يوجد لدينا قطع عرضي مكون من 7 قراءات لـ 7 دول. ويصبح لدينا عدد من المشاهدات المزدوجة عددها 175 مشاهدة (25 سنة x 7 دول).

بيانات القائمة، القطع الطولي، القوائم الصغيرة:

Panel, Longitudinal or micropanel data

كل هذه المسميات لنوع من البيانات المزدوجة، حيث تكون البيانات المزدوجة لنفس الوحدة (مثل نفس العائلة أو نفس المنشأة). فمثلاً قيام قسم U.S بإجراء تعداد (مسح) للسكان في فترات دورية. المعاينة الدورية لنفس السكان (الذين يعيشون في نفس العناوين)، حيث تكون المقابلة لتحديد هل طرأ أي تغير على حياة هؤلاء الأشخاص سواء من الناحية الاجتماعية أو المالية، . . الخ. والمقابلات بشكل دوري سوف تؤدي إلى توافر معلومات بشكل ديناميكي لسلوك هؤلاء السكان، وسوف نناقش ذلك بالتفصيل في الفصل (16).

مصادر البيانات : The source of data[12]

والبيانات المستخدمة في التحليل التجريبي (الاختباري) تقوم بجمعها جهات حكومية (مثل قسم التجارة) أوجهة دولية (مثل صندوق النقد الدولي أو البنك الدولي) أو المنظمات الخاصة (مثل هيئة رعاية الفقراء) كذلك يوجد آلاف من الهيئات الأخرى تقوم بجمع البيانات بهدف أو بآخر.

الإنترنت : The Internet

الإنترنت أحدث ثورة في جمع البيانات. فبالاتصال بشبكة الإنترنت وإدخال كلمة المفتاح (مثل معدلات التبادل مثلاً) يمكن الدخول على مصادر البيانات

(12) For an illuminating account, see Albert T. Somers, The U.S. Economy Demystified: What the Major Economic Statistics Mean And their Significance for Business, D.C. Heath, Lexington, Mass., 1985.

والحصول عليها. ملحق E يعطي بعض الزيارات المتكررة لبعض المواقع التي يمكن منها الحصول على البيانات بدون تكاليف تقريباً.

والبيانات التي يتم جمعها عن طريق هيئات مختلفة يمكن أن تكون بيانات تجريبية أو غير تجريبية.

والبيانات التجريبية غالباً يتم تجميعها في العلوم الطبيعية، فالمستشفى يمكن أن يرغب في جمع بيانات عن عوامل (متغيرات) معينة يمكن تثبيتها في مستويات معينة لتحديد تأثير بعض هذه العوامل على الظاهرة محل الاعتبار. فعلى سبيل المثال، إذا كان الباحث يرغب في دراسة تأثير السمنة (البدانة) على مستوى ضغط الدم. فالباحث يمكنه جمع البيانات عند تثبيت مستوى الأكل، التدخين، المشروبات الكحولية عند أفراد العينة لكي يقلل من تأثير هذه العوامل (المتغيرات) على مستوى ضغط الدم. في العلوم الاجتماعية، عادة تكون البيانات غير تجريبية، وذلك يرجع إلى طبيعة الظواهر، حيث لا تخضع هذه الظواهر لتحكم الباحث[15]. فعلى سبيل المثال بيانات GNP، البطالة، أسعار الأسهم، .. الخ لا تخضع لتحكم الباحث.

دقة البيانات: The accuracy data[14]

وبالرغم من وفرة البيانات المتاحة للباحث الاقتصادي، إلا أن جودة هذه البيانات غالباً لا تكون جيدة، وذلك يرجع إلى الأسباب التالية:

أولاً: بيانات العلوم الاجتماعية غالباً بيانات ذات طبيعة غير تجريبية في طبيعتها. لذلك يوجد إمكانية لأخطاء المشاهدة (الرصد).

ثانياً: يوجد أخطاء في التقريب، وتظهر هذه الأخطاء أيضاً في جمع البيانات التجريبية.

ثالثاً: في كثير من الحالات التي سيتم جمع البيانات باستخدام استمارة استقصاء كثير من المفردات لا تستجيب لاستفاء البيانات، فعادة يتم استيفاء حوالي

(13) In the social sciences too sometimes one can have a controlled experiment. An example is given in execise 1.6.

(14) For a critical review, see O. Morgenstern, The Accutracy of Economic Observations, 2d ed., Princeton University Press, Princeton, N. J., 1963.

40% فقط، وبالتالي يبنى التحليل على جزء فقط من البيانات، وبالتالي تكون العينة المختارة من البيانات عينة متحيزة selectivity bias sample. فضلاً عن ذلك، توجد مشكلة أخرى بالنسبة لاستمارة الاستقصاء، وهي وجود بعض الأسئلة لم يتم استيفاؤها وبصفة خاصة الأسئلة المتعلقة بالأمور المالية، مما يؤدي إلى وجود تحيز إضافي آخر.

رابعاً: استخدام طرق المعاينة المختلفة يؤدي إلى الحصول على نتائج مختلفة يصعب مقارنة نتائج العينات المختلفة معاً.

خامساً: البيانات الاقتصادية المتاحة بشكل عام، تكون في مستوى تجميعي. فمثلاً البيانات الكلية (مثل GNP، العمالة، التضخم، البطالة) تكون متاحة في الاقتصاد ككل أو في الغالب وفقاً للمناطق الجغرافية. هذه البيانات التجميعية لا تعطينا رؤية واضحة عن مفردات البحث.

سادساً: ونظراً لأن البيانات تعتبر خصوصية (سرية)، لذا يتم نشرها في شكل تجميعي، فعلى سبيل المثال، الـ TRS غير متاح قانونياً كشف بيانات عن بنود إيرادات الضرائب، ولكن المتاح هو تلخيص هذه البيانات.

لذلك يكون غير ممكن تحديد مقدار إنفاق الفرد على الرعاية الصحية من الأفراد ذوي مستوى معين للدخل عن طريق البيانات التجميعية.

وغالباً يفشل التحليل الكلي ليتناسب مع ديناميكية سلوك الوحدات الجزئية. كذلك نجد أن قسم التجارة المسئول عن إجراء مسح للأعمال كل 5 سنوات غير متاح له الكشف عن معلومات عن الإنتاج، التوظيف، الطاقة، الاستهلاك، البحث، .. الخ، وعلى مستوى المنشأة يكون من الصعب دراسة الاختلافات الداخلية لهذه المنشآت.

لجميع الأسباب السابق ذكرها، لابد أن يكون واضحًا تماماً للباحث أن جودة النتائج التي يتم التوصل إليها مرتبط تماماً بجودة البيانات المستخدمة.

لذلك في بعض الأبحاث، نجد أن نتائج الأبحاث غير مرضية، ويكون السبب ليس استخدام نموذج غير مناسب ولكن السبب يرجع إلى افتقار البيانات.

ملاحظة على قياس تدرج المتغيرات a note on the measurement scales of variables[15]

عادة يمكن تقسيم تدرج قياس المتغيرات إلى 4 فئات على النحو التالي: تدرج النسبة، تدرج الفترة، التدرج الترتيبي، التدرج الاسمي.

التدرج النسبي: Ratio scale

إذا اعتبرنا المتغير X يأخذ القيمتين X_2, X_1 فان النسبة X_2/X_1 أو الفرق $(X_1 - X_2)$ تمثل كميات ذات معنى. ويوجد ترتيب طبيعي (تصاعدي أو تنازلي) لهذه القيم بين التدرج، لذلك تكون المقارنات مثل $X_2 = X_1$ أو $X_2 - X_1$ ذي معنى. ومعظم المتغيرات الاقتصادية تنتمي إلى هذه الفئة.

تدرج الفترة: Interval scale

هو أن يأخذ المتغير قيمًا معينة داخل الفترة مثل الفترة من 1995 إلى 2000 يكون له مغزى.

التدرج الترتيبي: Ordinal scale

وهنا يمكن أن يأخذ قيمًا ترتيبية، مثال ذلك فئات الدخل (عال أو متوسط، منخفض).

التدرج الاسمي: Nominal scale

وتوجد متغيرات أخرى لفظية مثل النوع gender (ذكر، أنثى) أو الحالة الاجتماعية (متزوج، مطلق، أعزب).

ومما هو جدير بالذكر، أن أساليب الاقتصاد القياسي الممكن تطبيقها على متغير النسبة مثلاً قد لا تكون مناسبة للتطبيق لمتغيرات الفترة. لذلك لابد أن يؤخذ في الاعتبار قياس تدرج المتغير.

8-1 الملخص والاستنتاجات SUMMARY AND CONCLUSIONS :

1 - الفكرة الرئيسة في تحليل الانحدار هو وجود تابع ومتغير أو أكثر مفسر.

2 - هدف التحليل هو تقدير أو التنبؤ بمتوسط قيمة المتغير التابع عند قيم محددة للمتغيرات المفسرة.

(15) The following discussion relies heavily on Aris Spanos, Probability Theory and Statistical Inference: Econometric Modeling with Obsrvational Data, Cambridge University Press, New York, 1999, p. 24.

3 - عملياً، نجاح تحليل الانحدار يتوقف على ملاءمة البيانات، وهذا الفصل ناقش طبيعة ومصادر، ومحددات البيانات بشكل عام، وبشكل خاص في العلوم الاجتماعية.

4 - في أي بحث، لابد أن يكون واضحًا للباحث مصادر البيانات المستخدمة في التحليل، والتعريفات وجمع البيانات، وأي فجوة ثم تنعكس على التحليل والنتائج.

5 - يمكن للقارئ استخدام البيانات الموجودة في المرجع.

تمارين : EXERCISES

1.1 جدول (2.1) يعطي الرقم القياسي لسعر الاستهلاك (CPI) لـ 7 دول صناعية حيث 1982–1984= 100 كأساس للقياس.

(أ) من البيانات، احسب معدل تضخم لكل دولة[16].

جدول (2.1) CPI لـ 7 دول صناعية خلال الفترة 1973–1997

Year	Canada	France	Germany	Italy	Japan	U.K.	U.S.
1973	40.80000	34.60000	62.80000	20.60000	47.90000	27.90000	44.40000
1974	45.20000	39.30000	67.10000	24.60000	59.00000	32.30000	49.30000
1975	50.10000	43.90000	71.10000	28.80000	65.90000	40.20000	53.80000
1976	53.90000	48.10000	74.20000	33.60000	72.20000	46.80000	56.90000
1977	58.10000	52.70000	76.90000	40.10000	78.10000	54.20000	60.60000
1978	63.30000	57.50000	79.00000	45.10000	81.40000	58.70000	65.20000
1979	69.20000	63.60000	82.20000	52.10000	84.40000	66.60000	72.60000
1980	76.10000	72.30000	86.70000	63.20000	90.90000	78.50000	82.40000
1981	85.60000	81.90000	92.20000	75.40000	95.30000	87.90000	90.90000
1982	94.90000	91.70000	97.10000	87.70000	98.10000	95.40000	96.50000
1983	100.4000	100.4000	100.3000	100.8000	99.80000	99.80000	99.60000
1984	104.7000	108.1000	102.7000	111.5000	102.1000	104.8000	103.9000
1985	109.0000	114.4000	104.8000	121.1000	104.1000	111.1000	107.6000
1986	113.5000	117.3000	104.7000	128.5000	104.8000	114.9000	109.6000
1987	118.4000	121.1000	104.9000	134.4000	104.8000	119.7000	113.6000
1988	123.2000	124.4000	106.3000	141.1000	105.6000	125.6000	118.3000
1989	129.3000	128.7000	109.2000	150.4000	108.1000	135.3000	124.0000
1990	135.5000	133.0000	112.2000	159.6000	111.4000	148.2000	130.7000
1991	143.1000	137.2000	116.3000	169.8000	115.0000	156.9000	136.2000
1992	145.3000	140.5000	122.1000	178.8000	116.9000	162.7000	140.3000
1993	147.9000	143.5000	127.6000	186.4000	118.4000	165.3000	144.5000
1994	148.2000	145.8000	131.1000	193.7000	119.3000	169.4000	148.2000
1995	151.4000	148.4000	133.5000	204.1000	119.1000	175.1000	152.4000
1996	153.8000	151.4000	135.5000	212.0000	119.3000	179.4000	156.9000
1997	156.3000	153.2000	137.8000	215.7000	121.3000	185.0000	160.5000

(16) Subtract from the current year's CPI the CPI from the previous year, divide the difference by the previous year's CPI and multiply the result by 100. Thus, the inflation rate for Canada for 1974 is [(45.2–40.8)/40.8] x 100 = 10.78% (approx.)

(ب) ارسم معدل التضخم لكل دولة من الدول الـ 7 وفقاً للزمن (اعتبر المحور الأفقي يمثل الزمن، والمحور الرأسي يمثل معدل التضخم).

(جـ) ماهي أهم الاستنتاجات عن التضخم في الدول الـ 7.

(د) ماهي الدول التي يعتبر معدل التضخم فيها أكثر تغيراً، ثم عقب على ذلك.

2.1 (أ) ارسم معدل التضخم لكل من الدول: كندا، فرنسا، ألمانيا، إيطاليا، اليابان، المملكة المتحدة، في مقابل معدل التضخم في الولايات المتحدة.

(ب) عقب بشكل عام حول سلوك معدل التضخم في كل دولة من الدول الـ 6 السابقة بالمقارنة بمعدل التضخم في الولايات المتحدة.

(جـ) في حالة إذا كانت معدلات التضخم في الدول الـ 6 تتحرك في نفس اتجاه معدل التضخم في الولايات المتحدة، فهل هذا يعني أن التضخم في الولايات المتحدة يتسبب في التضخم في الدول الأخرى؟ ولماذا؟

3.1 جدول (3.1) يوضح معدلات التبادل الخارجية لـ 7 دول صناعية خلال الفترة 1977-1998 . باستثناء المملكة المتحدة، يعرف معدل التبادل بأنه عدد الوحدات من العملة الأجنبية التي تساوي دولاراً واحداً أمريكيًا، أما في المملكة المتحدة فيعرف معدل التبادل بأنه عدد الدولارات الأمريكية التي تساوي جنيهًا استرلينيًا.

(أ) ارسم معدلات التبادل وفقاً للزمن ، ثم عقب على السلوك العام لمعدلات التبادل خلال الفترة الزمنية.

(ب) يقال إن الدولار زادت قيمته، في حالة شرائه عدد وحدات أكثر من العملة الأجنبية، والعكس صحيح، يقال إن قيمة الدولار قلت في حالة شرائه عدد وحدات أقل من العملة الأجنبية.

ماهو السلوك العام للدولار الأمريكي خلال الفترة 1977-1998؟ ابحث في مراجع الاقتصاد الكلي أو الاقتصاد الدولي عن عوامل تحديد زيادة أو نقص سعر العملة الأجنبية.

جدول (3.1) معدلات التبادل للدول الـ 7 خلال 1977-1998

Year	Canada	France	Germany	Japan	Sweden	Switzerland	U.K.
1977	1.063300	4.916100	2.323600	268.6200	4.480200	2.406500	1.744900
1978	1.140500	4.509100	2.009700	210.3900	4.520700	1.790700	1.918400
1979	1.171300	4.256700	1.834300	219.0200	4.289300	1.664400	2.122400
1980	1.169300	4.225100	1.817500	226.6300	4.231000	1.677200	2.324600
1981	1.199000	5.439700	2.263200	220.6300	5.066000	1.967500	2.024300
1982	1.234400	6.579400	2.428100	249.0600	6.283900	2.032700	1.748000
1983	1.232500	7.620400	2.553900	237.5500	7.671800	2.100700	1.515900
1984	1.295200	8.735600	2.845500	237.4600	8.270800	2.350000	1.336800
1985	1.365900	8.980000	2.942000	238.4700	8.603200	2.455200	1.297400
1986	1.389600	6.925700	2.170500	168.3500	7.127300	1.797900	1.467700
1987	1.325900	6.012200	1.798100	144.6000	6.346900	1.491800	1.639800
1988	1.230600	5.959500	1.757000	128.1700	6.137000	1.464300	1.781300
1989	1.184200	6.380200	1.880800	138.0700	6.455900	1.636900	1.638200
1990	1.166800	5.446700	1.616600	145.0000	5.923100	1.390100	1.784100
1991	1.146000	5.646800	1.661000	134.5900	6.052100	1.435600	1.767400
1992	1.208500	5.293500	1.561800	126.7800	5.825800	1.406400	1.766300
1993	1.290200	5.666900	1.654500	111.0800	7.795600	1.478100	1.501600
1994	1.366400	5.545900	1.621600	102.1800	7.716100	1.366700	1.531900
1995	1.372500	4.986400	1.432100	93.96000	7.140600	1.181200	1.578500
1996	1.363800	5.115800	1.504900	108.7800	6.708200	1.236100	1.560700
1997	1.384900	5.839300	1.734800	121.0600	7.644600	1.451400	1.637600
1998	1.483600	5.899500	1.759700	130.9900	7.952200	1.450600	1.657300

4.1 بيانات الجدول التالي (4.1) تعطي عرض النقود M1 السابق عرضها في شكل (5.1) وضح الأسباب التي أدت إلى زيادة عرض النقود خلال الفترة الموضحة في الجدول.

جدول (4.1) عرض M1 المعدلة موسمياً خلال الفترة 1959 :01-1999 :09 (بالمليون دولار)

1959:01	138.8900	139.3900	139.7400	139.6900	140.6800	141.1700
1959:07	141.7000	141.9000	141.0100	140.4700	140.3800	139.9500
1960:01	139.9800	139.8700	139.7500	139.5600	139.6100	139.5800
1960:07	140.1800	141.3100	141.1800	140.9200	140.8600	140.6900
1961:01	141.0600	141.6000	141.8700	142.1300	142.6600	142.8800
1961:07	142.9200	143.4900	143.7800	144.1400	144.7600	145.2000
1962:01	145.2400	145.6600	145.9600	146.4000	146.8400	146.5800
1962:07	146.4600	146.5700	146.3000	146.7100	147.2900	147.8200
1963:01	148.2600	148.9000	149.1700	149.7000	150.3900	150.4300
1963:07	151.3400	151.7800	151.9800	152.5500	153.6500	153.2900
1964:01	153.7400	154.3100	154.4800	154.7700	155.3300	155.6200
1964.07	156.8000	157.8200	158.7500	159.2400	159.9600	160.3000
1965:01	160.7100	160.9400	161.4700	162.0300	161.7000	162.1900
1965:07	163.0500	163.6800	164.8500	165.9700	166.7100	167.8500
1966:01	169.0800	169.6200	170.5100	171.8100	171.3300	171.5700
1966:07	170.3100	170.8100	171.9700	171.1600	171.3800	172.0300
1967:01	171.8600	172.9900	174.8100	174.1700	175.6800	177.0200
1967:07	178.1300	179.7100	180.6800	181.6400	182.3800	183.2600
1968:01	184.3300	184.7100	185.4700	186.6000	187.9900	189.4200

1968:07	190.4900	191.8400	192.7400	194.0200	196.0200	197.410
1969:01	198.6900	199.3500	200.0200	200.7100	200.8100	201.270
1969:07	201.6600	201.7300	202.1000	202.9000	203.5700	203.880
1970:01	206.2200	205.0000	205.7500	206.7200	207.2200	207.540
1970:07	207.9800	209.9300	211.8000	212.8800	213.6600	214.410
1971:01	215.5400	217.4200	218.7700	220.0000	222.0200	223.450
1971:07	224.8500	225.5800	226.4700	227.1600	227.7600	228.320
1972:01	230.0900	232.3200	234.3000	235.5800	235.8900	236.620
1972:07	238.7900	240.9300	243.1800	245.0200	246.4100	249.250
1973:01	251.4700	252.1500	251.6700	252.7400	254.8900	256.690
1973:07	257.5400	257.7600	257.8600	259.0400	260.9800	262.880
1974:01	263.7600	265.3100	266.6800	267.2000	267.5600	268.440
1974:07	269.2700	270.1200	271.0500	272.3500	273.7100	274.200
1975:01	273.9000	275.0000	276.4200	276.1700	279.2000	282.430
1975:07	283.6800	284.1500	285.6900	285.3900	286.8300	287.070
1976:01	288.4200	290.7600	292.7000	294.6600	295.9300	296.160
1976:07	297.2000	299.0500	299.6700	302.0400	303.5900	306.250
1977:01	308.2600	311.5400	313.9400	316.0200	317.1900	318.710
1977:07	320.1900	322.2700	324.4800	326.4000	328.6400	330.870
1978:01	334.4000	335.3000	336.9600	339.9200	344.8600	346.800
1978:07	347.6300	349.6600	352.2600	353.3500	355.4100	357.280
1979:01	358.6000	359.9100	362.4500	368.0500	369.5900	373.340
1979:07	377.2100	378.8200	379.2800	380.8700	380.8100	381.770
1980:01	385.8500	389.7000	388.1300	383.4400	384.6000	389.460
1980:07	394.9100	400.0600	405.3600	409.0600	410.3700	408.060
1981:01	410.8300	414.3800	418.6900	427.0600	424.4300	425.500
1981:07	427.9000	427.8500	427.4600	428.4500	430.8800	436.170
1982:01	442.1300	441.4900	442.3700	446.7800	446.5300	447.890
1982:07	449.0900	452.4900	457.5000	464.5700	471.1200	474.300
1983:01	476.6800	483.8500	490.1800	492.7700	499.7800	504.350
1983:07	508.9600	511.6000	513.4100	517.2100	518.5300	520.790
1984:01	524.4000	526.9900	530.7800	534.0300	536.5900	540.5400
1984:07	542.1300	542.3900	543.8600	543.8700	547.3200	551.1900
1985:01	555.6600	562.4800	565.7400	569.5500	575.0700	583.1700
1985:07	590.8200	598.0600	604.4700	607.9100	611.8300	619.3600
1986:01	620.4000	624.1400	632.8100	640.3500	652.0100	661.5200
1986:07	672.2000	680.7700	688.5100	695.2600	705.2400	724.2800
1987:01	729.3400	729.8400	733.0100	743.3900	746.0000	743.7200
1987:07	744.9600	746.9600	748.6600	756.5000	752.8300	749.6800
1988:01	755.5500	757.0700	761.1800	767.5700	771.6800	779.1000
1988:07	783.4000	785.0800	784.8200	783.6300	784.4600	786.2600
1989:01	784.9200	783.4000	782.7400	778.8200	774.7900	774.2200
1989:07	779.7100	781.1400	782.2000	787.0500	787.9500	792.5700
1990:01	794.9300	797.6500	801.2500	806.2400	804.3600	810.3300
1990:07	811.8000	817.8500	821.8300	820.3000	822.0600	824.5600
1991:01	826.7300	832.4000	838.6200	842.7300	848.9600	858.3300
1991:07	862.9500	868.6500	871.5600	878.4000	887.9500	896.7000
1992:01	910.4900	925.1300	936.0000	943.8900	950.7800	954.7100
1992:07	964.6000	975.7100	988.8400	1004.340	1016.040	1024.450
1993:01	1030.900	1033.150	1037.990	1047.470	1066.220	1075.610
1993:07	1085.880	1095.560	1105.430	1113.800	1123.900	1129.310
1994:01	1132.200	1136.130	1139.910	1141.420	1142.850	1145.650
1994:07	1151.490	1151.390	1152.440	1150.410	1150.440	1149.750
1995:01	1150.640	1146.740	1146.520	1149.480	1144.650	1144.240
1995:07	1146.500	1146.100	1142.270	1136.430	1133.550	1126.730
1996:01	1122.580	1117.530	1122.590	1124.520	1116.300	1115.470
1996:07	1112.340	1102.180	1095.610	1082.560	1080.490	1081.340

1997:01	1080.520	1076.200	1072.420	1067.450	1063.370	1065.990
1997:07	1067.570	1072.080	1064.820	1062.060	1067.530	1074.870
1998:01	1073.810	1076.020	1080.650	1082.090	1078.170	1077.780
1998:07	1075.370	1072.210	1074.650	1080.400	1088.960	1093.350
1999:01	1091.000	1092.650	1102.010	1108.400	1104.750	1101.110
1999:07	1099.530	1102.400	1093.460			

5.1 بافتراض قيامك بتطوير نموذج اقتصادي للأنشطة الجنائية، مثل عدد الساعات التي تم قضاؤها في الأنشطة الجنائية (مثل بيع الحبوب المحرمة). ماهي المتغيرات التي يجب أخذها في الاعتبار في النموذج، انظر مرجع Gary Becher[17].

6.1 الجدول التالي (5.1) المنشور في 1 مارس 1984 في مقالة في جريدة the wall street journal. يتضمن الجدول ميزانية الإعلان (بالمليون دولار) لـ 21 مؤسسة في 1983 والآثار (بالمليـون دولار) أسبوعياً عن طريق الرجوع إلى منتجات هذه المؤسسات. البيانات مبنية على عينة من 4000 من البالغين من المستخدمين لمنتجات هذه المؤسسات.

جدول (5.1) تأثير الإنفاق على الإعلان

Firm	Impressions, millions	Expenditure, millions of 1983 dollars
1. Miller Lite	32.1	50.1
2. Pepsi	99.6	74.1
3. Stroh's	11.7	19.3
4. Fed'l Express	21.9	22.9
5. Burger King	60.8	82.4
6. Coca Cola	78.6	40.1
7. McDonald's	92.4	185.9
8. MCI	50.7	26.9
9. Diet Cola	21.4	20.4
10. Ford	40.1	166.2
11. Levi's	40.8	27.0
12. Bud Lite	10.4	45.6
13. ATT/Bell	88.9	154.9
14. Calvin Klein	12.0	5.0
15. Wendy's	29.2	49.7
16. Polaroid	38.0	26.9
17. Shasta	10.0	5.7
18. Meow Mix	12.3	7.6
19. Oscar Meyer	23.4	9.2
20. Crest	71.1	32.4
21. Kibbles 'N Bits	4.4	6.1

(17) G. S. Becker, "Crime and Punishment: An Economic Approach," Journal of Political Economy, vol. 76, 1968, pp. 169 - 217.

(أ) ارسم التـأثيـرات على المحـور الرأسي والمنفق على الإعـلان على المحـور الأفقي.

(ب) ماذا يمكن استنتاجه حول طبيعة العلاقة بين المتغيرات في الجدول.

(جـ) وضح بيانياً كيف يمكن التفكير في الدفع للإعلان.

الفصل الثاني

تحليل انحدار المتغير.. بعض الأفكار الأساسية

VARIABLE REGRESSION ANALYSIS.. SOME BASIC IDEAS

في الفصل الأول، تناولنا مفهوم الانحدار بشكل عام. وفي هذا الفصل، سوف نتناول هذا الموضوع تفصيلياً. وبصفة خاصة، سوف نتناول ذلك في هذا الفصل، والفصلين التاليين. حيث يتم تناول النظريات الأساسية لتحليل الانحدار في حالة متغيرين فقط، أحدهما المتغير التابع، ومتغير مستقل واحد (مفسر). وتحليل الانحدار في متغيرين يعتبر الحالة البسيطة التي يتم تعميمها فيما بعد في حالة متغير واحد تابع، وأكثر من متغير مستقل (مفسر)، أو مايسمى بتحليل الانحدار المتعدد.

1.2 مثال افتراضي: A HYPOTHETICAL EXAMPLE[1]

وكما ذكرنا سابقاً في الفصل (2.1) أن تحليل الانحدار يهتم بتقدير أو التنبؤ بالقيمة المتوقعة للمتغير التابع عند قيم محددة للمتغير (أو المتغيرات) المفسر[2]. ولتوضيح ذلك، سوف نعتبر البيانات في جدول (1.2).

(1) The reader whose statistical knowledge has become somewhat rusty may want to freshen it up by reading the statistical appendix, App. A, before reading this chapter.

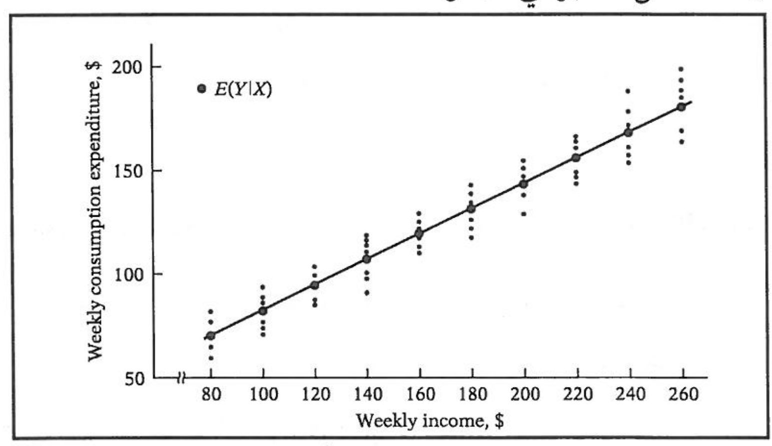
(2) The expected value, or expectation, or population mean of a random variable Y is denoted by the symbol $E(Y)$ On the other hand, the meam value computed f om a sample of values from the Y population is denoted as $\overline{Y}$, read as Y bar.

جدول (1.2) دخل الأسرة الأسبوعي X

$Y\downarrow$ \ $X\rightarrow$	80	100	120	140	160	180	200	220	240	260
Weekly family consumption expenditure Y, $	55	65	79	80	102	110	120	135	137	150
	60	70	84	93	107	115	136	137	145	152
	65	74	90	95	110	120	140	140	155	175
	70	80	94	103	116	130	144	152	165	178
	75	85	98	108	118	135	145	157	175	180
	–	88	–	113	125	140	–	160	189	185
	–	–	–	115	–	–	–	162	–	191
Total	325	462	445	707	678	750	685	1043	966	1211
Conditional means of Y, $E(Y\mid X)$	65	77	89	101	113	125	137	149	161	173

ويوضح الجدول توزيع الإنفاق الاستهلاكي الأسبوعي (Y) بالدولار للأسرة وفقاً لمستويات الدخل الأسبوعي (X) بالدولار، وذلك لعدد 60 أسرة تمثل مجتمع الدراسة population .

حيث تم توزيع 60 أسرة على 10 مجموعات للدخل الأسبوعي (من 80$ إلى 260$) وفقاً للدخل الأسبوعي للأسرة .



شكل (1.2) التوزيع الشرطي للإنفاق وفقاً لمستويات الدخل المختلفة

ومن الشكل، يتضح أن متوسط الإنفاق الاستهلاكي يتزايد بتزايد مستوى الدخل. ومن جدول (1.2) يتضح متوسط الإنفاق الاستهلاكي الأسبوعي للأسرة عند كل مستوى من مستويات الدخل. فعلى سبيل المثال، نجد أن متوسط الإنفاق

الاستهلاكي للأسرة عند الدخل 80 $ يساوي 65 $، كذلك عند الدخل 100 $ يساوي 77 $. وكل متوسط عند دخل معين يسمى بالقيمة المتوقعة الشرطية للإنفاق عند دخل معين ويشار إليها بالرمز $E(Y|X)$ وتقرأ التوقع الشرطي للمتغير Y بشرط X، وبالتالي فإن $E(Y|X=80) = 65$، كذلك $E(Y|X=100) = 77$ وأنه من الأهمية توضيح الفرق بين التوقع الشرطي لـY بشرط X والمشار إليه بالرمز $E(Y|X)$ وبين التوقع غير الشرطي لـY يشار إليه بالرمز $E(Y)$ وتقرأ "توقع Y".

التوقع غير الشرطي لـY يمكن حسابه على النحو التالي:

$$E(Y) = \frac{\sum Y}{60} = \frac{7272}{60} = \$121.20$$ [3]

وللتمييز بين القيمة المتوقعة غير المشروطة $E(Y)$ والقيمة المتوقعة المشروطة $E(Y|X)$، فالقيمة المتوقعة للإنفاق الاستهلاكي الأسبوعي للأسرة تساوي $E(Y)$ أي $E(Y) = 121.20$، ولكن القيمة المتوقعة للإنفاق الاستهلاكي الأسبوعي لأسرة بشرط أن يكون دخلها دخلاً معيناً وليكن القيمة المتوقعة لإنفاق الأسرة التي دخلها 140 $ تساوي 101 $ أي أن:

$$E(Y|X=140) = \$101$$

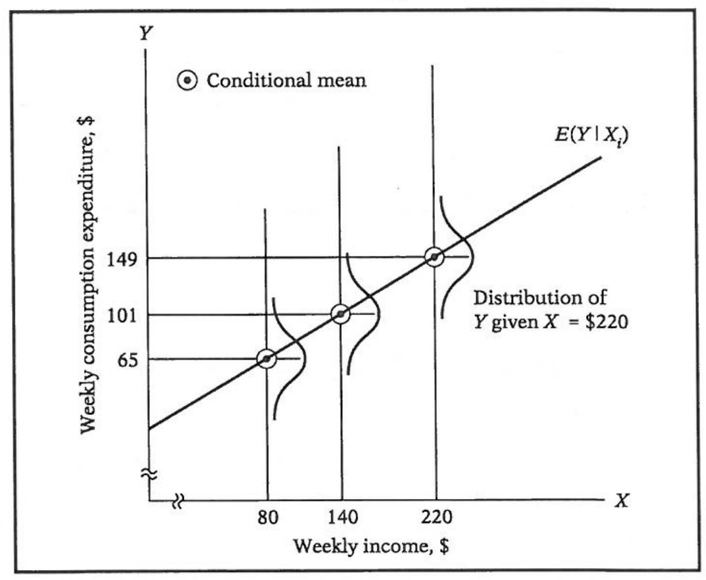
جدول (2.2) الاحتمالات الشرطية $p(Y|X)$ لبيانات جدول (1.2)

$p(Y \mid X_i)$ ↓ / $X \rightarrow$	80	100	120	140	160	180	200	220	240	260
Conditional probabilities $p(Y \mid X_i)$	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	$\frac{1}{5}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{5}$	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	–	$\frac{1}{6}$	–	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{6}$	–	$\frac{1}{7}$	$\frac{1}{6}$	$\frac{1}{7}$
	–	–	–	$\frac{1}{7}$	–	–	–	$\frac{1}{7}$	–	$\frac{1}{7}$
Conditional means of Y	65	77	89	101	113	125	137	149	161	173

(3) As shown in **App. A,** in general the conditional and unconditional mean values are different.

هكذا معرفة مستوى الدخل (X) مسبقاً يمكننا من معرفة[4] القيمة المتوقعة للإنفاق الاستهلاكي المشروط بالمستوى X ويأخذ القيمة $E(Y|X)$. وفي شكل (2-1) بتوصيل النقط التي تشير إلى $E(Y|X)$ عند قيم X مختلفة فنجد أنها تقع على خط واحد. ويسمى الخط بخط انحدار المجتمع population regression line (PRL)، وأكثر تعميماً يسمى منحنى انحدار المجتمع population regression curve[5] وللاختصار يسمى انحدار Y على X.

وهندسياً، فإن منحنى المجتمع هو عبارة عن المنحنى الذي تقع عليه النقط التي تمثل القيمة المتوقعة المشروطة للمتغير Y عند قيم معينة للمتغير X أي المنحنى الواصل بين النقط $E(Y|X)$ المختلفة كما هو موضح بشكل (2.2).



شكل (2.2) يوضح خط انحدار المجتمع (لبيانات جدول 1.2)

شكل (2.2) يوضح أنه عند كل قيمة لـ X (أي عند كل مستوى للدخل) توجد قيم متعددة للمتغير Y (أي قيم متعددة للإنفاق الاستهلاكي الأسبوعي). وهذه القيم المتعددة للمتغير Y عند قيمة معينة للمتغير X، تكون منتشرة حول قيمة التوقع الشرطي، أي حول $E(Y|X)$.

---

(4) I am indebted to James Davidson on this perspective. See James Davidson, Econometric Theory, Blackwell Publishers, Oxford, U.K., 2000, p. 11.

(5) In the present example the PRL is a straight line, but it could be a curve (see Figure 2.3).

وللتبسيط، يفترض أن هذه القيم المنتشرة لـ Y تتوزع توزيعاً متماثلاً symmetric حول قيمة التوقع الشرطي.

2.2 مفهوم دالة انحدار المجتمع :
THE CONCEPT OF POPULATION REGRESSION FUNCTION (PRF)

من العرض السابق لشكلي (1.2، 2.2) يتضح أن التوقع الشرطي $E(Y|X_i)$ يمثل دالة في X_i حيث X_i قيمة معينة للمتغير X أوبعبارة أخرى

$$E(Y \mid X_i) = f(X_i) \qquad (1.2.2)$$

حيث $f(X_i)$ تشير إلى دالة ما في المتغير المفسر X. وفي المثال المعطى سابقاً، فإن $E(Y|X_i)$ يمثل دالة خطية في X_i. والمعادلة (1.2.2) تعرف باسم معادلة التوقع الشرطي أو دالة انحدار المجتمع أو انحدار المجتمع للاختصار وهذا يعني أن القيمة المتوقعة للمتغير Y عند قيمة معينة X_i مرتبطة دالياً بـ X_i.

ويصبح السؤال الآن هو: ماهو شكل الدالة $f(X_i)$ ؟ وهذا يعتبر سؤالاً ذا أهمية، لأنه في المشاكل الفعلية لايكون لدينا بيانات كاملة عن المجتمع ككل، بالإضافة إلى افتراض أن شكل الدالة يتوقف على عوامل متعددة.

ولكن نظرياً يمكن على سبيل المثال افتراض العلاقة الخطية بين الدخل X_i والإنفاق الاستهلاكي Y على النحو التالي:

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i. \qquad (2.2.2)$$

حيث β_1 ، β_2 مقادير ثابتة غير معلومة تسمى بمعاملات الانحدار والمعادلة (1.2.2) تعتبر دالة انحدار المجتمع الخطية. وأحياناً يعبر عنها بنموذج انحدار المجتمع الخطي أو انحدار المجتمع الخطي البسيط. وبصفة عامة تستخدم التعبيرات التالية لمرادفات نفس الشئ: الانحدار، معادلة الانحدار، نموذج الانحدار.

وتحليل الانحدار يهتم بتقدير الدالة في (2.2.2) أي تقدير قيم β_1 ، β_2 من خلال المشاهدات عن قيم x، وقيم y المناظرة لها. وسوف نتناول ذلك بالتفصيل في الفصل الثالث.

3.2 معنى مصطلح الخطي :

THE MEANING OF THE TERM LINEAR

ويهتم هذا الكتاب أساساً بالنماذج الخطية كما في (2.2.2). ومن الأهمية معرفة أنه يمكن تفسير النماذج الخطية بطريقتين مختلفتين، كما سوف نوضح فيما يلي.

الخطية في المتغيرات : linearity in the variables

وهنا يعني أن التوقع الشرطي $E(Y|X_i)$ دالة خطية في المتغير X_i كما هو في (2.2.2)[6]. وفي هذه الحالة يكون منحنى الانحدار عبارة عن خط مستقيم. أما إذا كان التوقع الشرطي على النحو:

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_2^i$$

في هذه الحالة لا تكون الدالة خطية في X_i.

الخطية في المعلمات : linearity in the parameters

والتفسير الثاني للخطية وهو أن الدالة $E(Y \mid X_i)$ دالة خطية في المعلمات β's وليس في المتغير X_i [7].

فمثلاً الدالة :

$$E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$$

تعتبر دالة خطية في المعلمات β_1, β_2 فمثلاً:

$$E(Y \mid X = 3) = \beta_1 + 9\beta_2$$

فهي دالة خطية في β_1، β_2.

وجميع النماذج في شكل (3.2) تعتبر نماذج خطية في المعلمات. أما النموذج:

$$E(Y \mid X_i) = \beta_1 + \beta_2^2 X_i$$

---

(6) A function $Y = f(X)$ is said to be linear in X if X appears with a power or index of 1 only (that is, terms such as $X^2, \sqrt{X}$ and so on, are excluded) and is not multiplied or divided by any other variable (for example, $X \cdot Z$ or X/Z, where Z is another variable). If Y depends on X alone, anotherway to state that Y is linearly related to X is that the mte of change of Y with respect to X (i.e., the slope, or derivative of Y with respect to X, dY/dX) is independent of the value of X Thus if $Y = 4X$ $dY/dX = 4$ which is independent of the value of X But if $Y = 4X^2$ $dY/dX = 8X$ which is not independent of the value ta ken by X. Hence this function is not linear in X

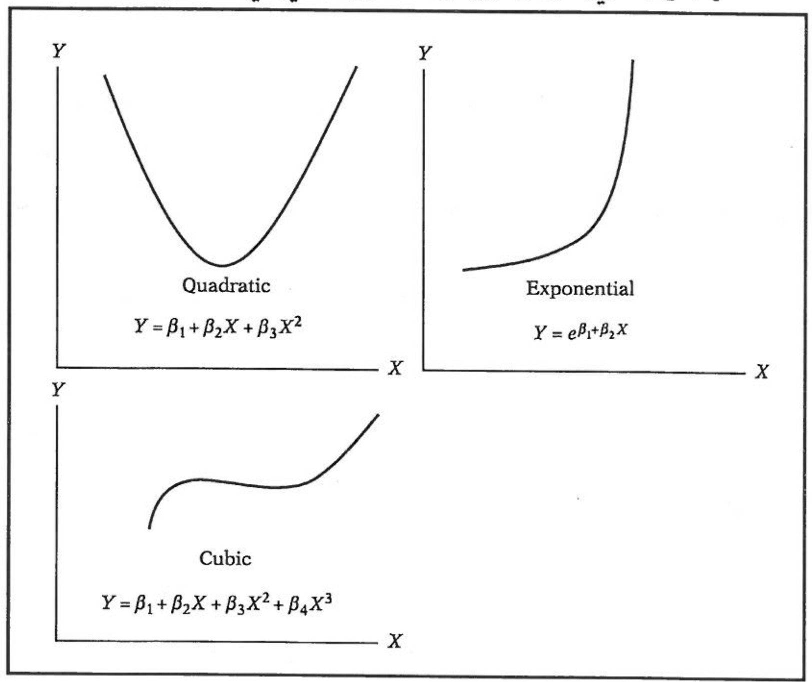
(7) A function is said to be linear in the pammeter, say, β_1, if β_1 appears with a power of 1 only amd is not multiplied or divided by any other parameter (for example, $\beta_1\beta_2$, β_2/β_1, and so on).

وعند X_i=3 نجد أن :

$$E(Y \mid X_i = 3) = \beta_1 + 3\,\beta_2^2$$

يعتبر نموذجاً غير خطي في المعلمات، وسوف نناقش هذا النوع من النماذج في الفصل (14). وسوف نتناول التفسيرين للخطية في هذا الكتاب. ومن الآن سوف نعتبر الخطية للانحدار خطية في المعلمات، في حين ممكن أن تكون المتغيرات المفسرة X's خطية أو غير خطية.

هكذا يعتبر النموذج $E(Y \mid X_i) = \beta_1 + \beta_2 X_i$ نموذجاً خطياً في المعلمات β_1، β_2، كذلك خطي في المتغير المفسر X_i. أما النموذج $E(Y \mid X_i) = \beta_1 + \beta_2 X_i^2$ نموذج خطي في المعلمات β_1، β_2، في حين يعتبر نموذجاً غير خطي في المتغير X_i.



شكل (3.2) دوال خطية في المعلمات

جدول (3.2) نماذج الانحدار الخطية

Model linear in parameters?	Model linear in variables?	
	Yes	No
Yes	LRM	LRM
No	NLRM	NLRM

Note: LRM = linear regression model
NLRM = nonlinear regression model

4.2 التحديد العشوائي لدالة انحدار المجتمع:
STOCHASTIC SPECIFICATION OF PRF

ويتضح من شكل (1.2) أنه كلما زاد دخل الأسرة زاد أيضاً متوسط الإنفاق الاستهلاكي. ولكن ما هي علاقة استهلاك الأسر عند مستوى محدد من الدخل s في جدول (1.2) كذلك شكل (1.2) يتضح أنه ليس بالضرورة زيادة الإنفاق الاستهلاكي عند زيادة مستوى الدخل. فعلى سبيل المثال، من جدول (1.2) نجد أنه وفقاً لمستوى الدخل 100 $ توجد أسرة تنفق 65 $ أي أقل من إنفاق استهلاك أسرتين دخل كل منهما 80 $.

ولكن نلاحظ أن متوسط إنفاق الأسر بدخل كل منهم 100 $ أكبر من متوسط الإنفاق الاستهلاكي لأسر الدخل الأسبوعي لكل منهم 80 $ (77 $ مقابل 65 $). ومن هنا يطرح السؤال التالي: ماهي العلاقة بين الإنفاق الاستهلاكي للأسر عند مستوى دخل معين من الدخل؟. من شكل (2-1) يتضح أنه عند مستوى دخل معين وليكن X_i الإنفاق الاستهلاكي لكثير من الأسر يتمركز عند متوسط الإنفاق الاستهلاكي بشرط الدخل X_i أي يتركز عند $E(Y|X_i)$. لذلك يمكن التعبير عن انحراف deviation الإنفاق الاستهلاكي Y_i عند متوسط توقع الإنفاق الاستهلاكي بشرط X_i على النحو التالي:

$$u_i = Y_i - E(Y|X_i)$$

$$Y_i = E(Y|X_i) + u_i \qquad (1.4.2)$$

حيث u_i تشير إلى انحراف Y_i عن $E(Y|X_i)$. وتمثل u_i متغيراً عشوائياً random variable ممكنًا أن يأخذ قيماً سالبة أو موجبة أو صفرية. ويسمى u_i بالاختلاف العشوائي stochastic disturbance أو الخطأ العشوائي stochastic error term.

ويعتبر الإنفاق الاستهلاكي Y_i عبارة عن جزءين: الجزء الأول عبارة عن $E(Y|X_i)$ وهو جزء يقيني (غير عشوائي) deterministic، والجزء الثاني هو u_i وهو عبارة عن جزء عشوائي random. وإذا فرضنا أن $E(Y|X_i)$ تأخذ الشكل الخطي كما في (2.2.2)، وبالتالي المعادلة (2.4.1) يمكن إعادة كتابتها على النحو:

$$\begin{aligned} Y_i &= E(Y|X_i) + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned} \qquad (2.4.2)$$

والمعادلة (2.4.2) يمكن كتابتها بالتفصيل عند $X_i = \$80$ على النحو التالي:

$$\begin{aligned} Y_1 &= 55 = \beta_1 + \beta_2(80) + u_1 \\ Y_2 &= 60 = \beta_1 + \beta_2(80) + u_2 \\ Y_3 &= 65 = \beta_1 + \beta_2(80) + u_3 \\ Y_4 &= 70 = \beta_1 + \beta_2(80) + u_4 \\ Y_5 &= 75 = \beta_1 + \beta_2(80) + u_5 \end{aligned} \quad (3.4.2)$$

وإذا أخذنا القيمة المتوقعة لطرفي المعادلة (1.4.2) نحصل على التالي:

$$\begin{aligned} E(Y_i \mid X_i) &= E[E(Y \mid X_i)] + E(u_i \mid X_i) \\ &= E(Y \mid X_i) + E(u_i \mid X_i) \end{aligned} \quad (4.4.2)$$

حيث يلاحظ أن توقع المقدار الثابت يساوي نفس المقدار الثابت[8].

وبما أن $E(Y_i \mid X_i)$ يساوي $E(Y \mid X_i)$ في المعادلة (4.4.2) وبالتالي فإن:

$$E(u_i \mid X_i) = 0 \quad (5.4.2)$$

هكذا افتراض أن خط الانحدار يمر بنقط التوقعات الشرطية للمتغير Y عند القيم المختلفة لـ X_i أي يمر بالنقط $E(Y \mid X_i)$، $i = 1, 2, ..., m$ (كما في شكل 2.2) فهذا يؤدي إلى أن التوقع الشرطي لـ u_i أي $E(u_i \mid X_i)$ يساوي الصفر لجميع قيم $i= 1, 2, ..., m$. ومما سبق يتضح أن العلاقة (2.2.2) تكافيء (2.4.2) إذا كان $E(u_i \mid X_i)= 0$ [9].

والتحديد العشوائي (2.4.2) stochastic specification يتميز بتوضيح أنه يوجد متغيرات أخرى غير مستوى الدخل تؤثر على مستوى الإنفاق الاستهلاكي. وأن مستوى الإنفاق الاستهلاكي يمكن التعبير عنه فقط باستخدام المتغيرات المتضمنة في نموذج الانحدار.

5.2 معنوية حد الاختلاف العشوائي : THE SIGNIFICANCE OF THE STOCHASTIC DISTURBANCE TERM

ومن الشكل (2-4) يتضح أن حد الاختلاف أو الخطأ العشوائي u_i يمثل مجموع المتغيرات المؤثرة على المتغير التابع Y، وغير موجودة بالنموذج وتأثيرها تأثيراً تجميعياً collectivety affect على المتغير التابع Y. ومن هنا يظهر السؤال التالي: لماذا لم يتضمن

(8) See **App. A** for a brief discussion of the properties of the expectation opemtor E. Note that $E(Y| X_i)$, once the value of X_i is fixed, is a constant.

(9) As a matter of fact, in the method of least squares to be developed in Chap. 3, it is assumed explicit Y that $E(u_i \mid X_i) = 0$. See Sec. 3.2.

النموذج هذه المتغيرات، ويصبح النموذج نموذج انحدار متعدد . وإجابة هذا السؤال ترجع إلى سبب أو أكثر من الأسباب التالية :

1 – الغموض النظري: غالباً تحديد سلوك المتغير Y يكون غير كامل. فنحن نعلم أن مستوى الدخل X له تأثير على الإنفاق الاستهلاكي Y، ولكن توجد متغيرات أخرى مؤثرة أيضاً في الإنفاق الاستهلاكي (مثل مستوى تعليم رب الأسرة، عدد أفراد الأسرة، . . الخ). لذلك يمكن استخدام المتغير u_i للتعبير عن تأثير هذه المتغيرات غير الموجودة بالنموذج.

2 – عدم إتاحة البيانات: غالباً يكون غير متاح البيانات الكمية لهذه المتغيرات التي يعبر عن تأثيرها التجميعي بالمتغير u_i. فعلى سبيل المثال، أساساً يتم اعتبار ثروة الأسرة هي المتغير المفسر X_2 آخر بجانب X_1 دخل الأسرة. ولكن عادة لاتوجد بيانات متاحة عن ثروة الأسرة. لذلك يتم اعتبار دخل الأسرة كمتغير مفسر فقط، ويتضمن المتغير العشوائي u_i تأثير ثروة الأسرة.

3 – المتغيرات الجوهرية مقابل المتغيرات السطحية : نجد أن الإنفاق الاستهلاكي Y يعتمد على مستوى الدخل X_1 كذلك يتأثر بعدد الأطفال في الأسرة X_2، والنوع X_1، والديانة X_4، . . . ولكن نجد أن أهم المتغيرات المفسرة هي X_1، في حين أن X_2، X_2، . . يكون تأثيرها أقل، بالإضافة إلى أن تكاليف تجميع بيانات عن هذه المتغيرات X2، X3، . . . إن أمكن تكون مكلفة، ولكن يمكن أخذ تأثيرها التجميعي في متغير u_i[10].

4 – الجوهر العشوائي في السلوك الإنساني intrinsic randomness in human behavior : وعادة السلوك الإنساني توجد فيه العشوائية randomness، فحتى إذا تمت المتغيرات الأخرى غير مستوى الدخل في النموذج، فإنه يظل وجود جزء عشوائي يتضمنه المتغير التابع Y لذا يعبر عنه في u_i.

5 – فقر المتغيرات التعويضية poor proxy variables : النموذج التقليدي للانحدار (الذي سوف يقيم في الفصل 3) يفترض أن المتغير Y والمتغير X مقاس كل منهما

(10) A further difficulty is that variables such as sex, education, and religion are difficult to quantify.

بدقة، ولكن عملياً البيانات يوجد بها بعض الأخطاء في القياس. وسوف تعتبر على سبيل المثال النظرية المعروفة لـ Milton Friedman's المعروفة بدالة الاستهلاك[11]. هو اعتبر الاستهلاك المستمر permanent consumption (Y^p) كدالة في الدخل المستمر permanent income(X^p). ولكن عملياً بيانات المتغيرين X، Y لايتم جمعهما ورصدهما بشكل مستمر. وبالتالي أصبح عملياً المتغيران X، Y هما بديلان عن X^p، Y^p وبالتالي يؤدي ذلك إلى وجود بعض الاختلافات والأخطاء يمكن أن يمثل تأثيرها في المتغير العشوائي u_i ويتم انعكاس ذلك في تقدير قيم المعلمات β.

6 – مبدأ التقطير principle of parsimony: بالرجوع إلى مرجع occam's razor[12] نجد أنه دائماً يفضل بناء نموذج انحدار بسيط بقدر الإمكان.

7 – الصياغة الدالية الخاطئة wrong functional form: حتى إذا كانت المتغيرات التابعة والمفسرة صحيحة نظرياً، وحتى لو توافرت البيانات المطلوبة لبناء النموذج، ولكن تظل مشكلة ماهي الصيغة الرياضية لشكل النموذج. هل النموذج خطي على النحو:

$$Y_i = \beta_1 + B_2X_i + u_i$$

أما نموذج غير خطي على النحو:

$$Y_i = \beta_1 + \beta_2X_i + \beta_3X_i^2 + u_i$$

أم شكل آخر. ويمكن تحديد الشكل الدالي للنموذج في حالة وجود متغير تابع Y، ومتغير مفسر واحد من خلال شكل الانتشار scattergram ولكن في حالة الانحدار المتعدد، أي وجود أكثر من متغير مفسر، فإنه في هذه الحالة، لايمكن استخدام شكل الانتشار بشكل مباشر: لجميع الأسباب السابقة يصبح افتراض المتغير العشوائي u_i يلعب دوراً مهماً في تحليل الانحدار.

(11) Milton Friedmam, A Theory of the Consumption Function, Princeton University Press, Princeton, N.J., 1957.

(12) "That descriptions be kept as simple as possible until proved inadequate," The World of Mathematics, vol. 2, J. R. Newman (ed.), Simon & Schuster, New York 1956, p. 1247, or, "Endties should not be multiplied beyond necessity," Donald F. Morrison, Applied Linear Statistical Methods, Prentice Hal1, Englewood Cliffs, N.J., 1983, p. 58.

6.2 دالة انحدار العينة :
THE SAMPLE REGRESSION FUNCTION (SRF)

في مناقشتنا السابقة، كان تعاملنا مع المجتمع وبيانات المجتمع محل الدراسة، وذلك تجنباً لمشاكل العينات (حيث تمثل بيانات جدول (1.2) بيانات مجتمع وليس عينة). ولكن في معظم الدراسات العملية نتعامل مع العينة. لذلك عملنا حالياً سوف ينحصر في كيفية تقدير PRF على أساس معلومات عينة Sample information. وسوف نلاحظ أن بيانات جدول (1.2) تمثل بيانات عينة.

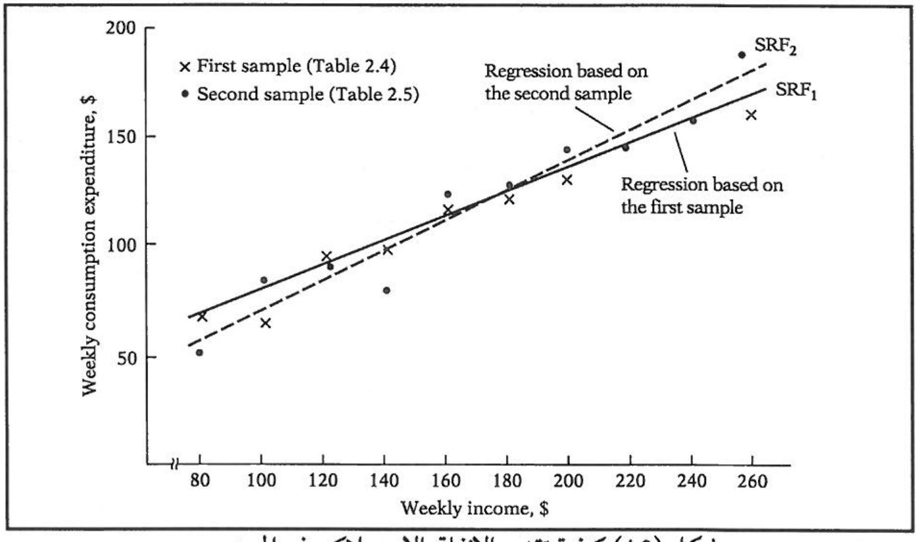
والسؤال الآن هو كيف يمكن التنبؤ (تقدير) الإنفاق الاستهلاكي Y في المجتمع عند قيمة معينة لـX؟. أو بعبارة أخرى تقدير PRF باستخدام بيانات عينة؟ ولتوضيح ذلك دعنا نرسم بيانات جدولي (4.2، 5.2) في شكل انتشار واحد في شكل (4.2).

جدول (2-4)

Y	X
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

جدول (2-5)

Y	X
55	80
88	100
90	120
80	140
118	160
120	180
145	200
135	220
145	240
175	260



شكل (4.2) كيفية تقدير الإنفاق الاستهلاكي في المجتمع

لتوفيق fit أشكال انتشار ملائمة: دالة انحدار العينة SRF لبيانات العينة الأولى في جدول (4.2)، كذلك دالة انحدار العينة الثانية SRF_2 المبنية على بيانات العينة الثانية بجدول (5.2) كما هو موضح بشكل (4.2). وتعتبر كل دالة من الدوال SRF_1 ، SRF_2 تقديراً لدالة انحدار المجتمع PRF المسحوب منه العينات. وكما أوضحنا سابقاً دالة انحدار المجتمع PRF (في العلاقة (2.2.2))، يمكن تطوير مفهوم دالة الانحدار للعينة SRF على النحو التالي:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad (1.6.2)$$

حيث $\hat{Y}$ تشير إلى تقدير (متوسط) قيمة Y عند X_i وتقرأ "Y هات" حيث $\hat{Y}$ هي تقدير لـ $E(Y \mid X_i)$ كذلك $\hat{\beta}_1$ ، $\hat{\beta}_2$ هي تقديرات لكل من β_1، β_2 على الترتيب وتسمى التقديرات $\hat{\beta}_2$ ، $\hat{\beta}_1$ المحسوبة من بيانات العينة بالإحصاءات statistics.

وكما سبق أن عبرنا عن PRF في صياغتين متكافئتين في (2.2.2)، (2.4.2)، أنه سوف نعبر عن SRF في (1.6.2) في صياغة عشوائية على النحو التالي:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (2.6.2)$$

حيث تشير u_i إلى حد الباقي residual term في العينة أو بعبارة أخرى هو تقدير لـ u_i في المجتمع.

ومما هو جدير بالذكر، أن هدفنا الأساسي هو تقدير دالة الانحدار للمجتمع population regression function PRF.

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (2.4.2)$$

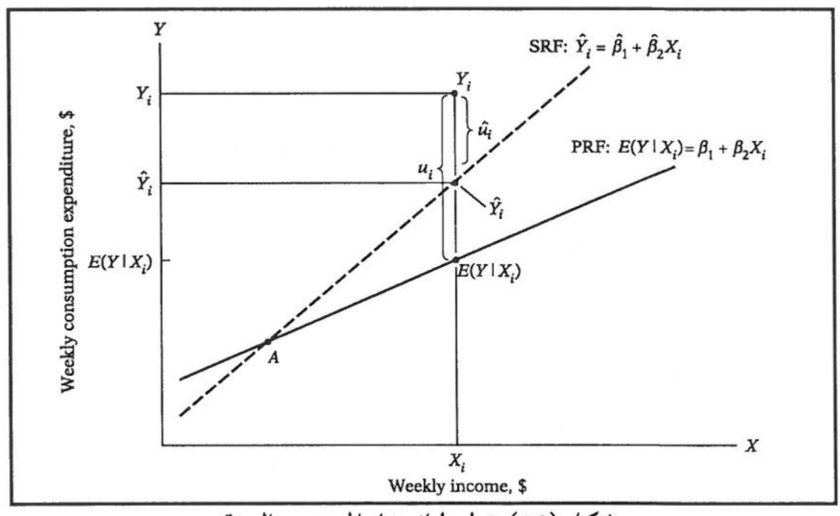
على أساس بناء دالة الانحدار على أساس بيانات العينة SRF تصبح على النحو التالي:

$$Y_i = \hat{\beta}_1 + \hat{\beta} x_i = \hat{u}_i \quad (2.6.2)$$

ولكن بسبب تأثير تقلبات المعاينة على تقدير PRF على أساس SRF.

ونظراً لأن التحليل يتم بناء على بيانات عينة واحدة من المجتمع.

(13) As noted in the Introduction, a hat above a variable will signify an estimator of the relevant population value.



شكل (5.2) خطوط انحدار المجتمع والعينة

وعادة الدالة SRF المستخدمة كتقدير للدالة PRF تكون أفضل تقريب للدالة PRF. ففي شكل (5.2) عندما $X = X_i$ نجد أنه من بيانات العينة عندما $Y = Y_i$ نجد أن:

$$Y_i = \hat{Y}_i + \hat{u}_i \quad (3.6.2)$$

كذلك نجد باستخدام الدالة PRF أن:

$$Y_i = E(Y | X_i) + u_i \quad (4.6.2)$$

ففي شكل (5.2) نجد أن $\hat{Y}_i$ كتقدير لـ $E(Y | X_i)$ تأخذ قيمة أكبر من $E(Y | X_i)$ تأخذ قيمة أقل من $E(Y | X_i)$ عند X_i، وعند أي نقطة يسار النقطة A نجد أن $\hat{Y}_i$ كتقدير لـ $E(Y | X_i)$ تأخذ قيمة أقل من $E(Y | X_i)$.

والسؤال الآن، هو كيف يمكن إيجاد SRF بحيث تكون أقرب مايمكن للدالة PRF أو بعبارة أخرى كيف يمكن إيجاد $\hat{\beta}_1$ ، $\hat{\beta}_2$ حيث يكونان أقرب ما يمكن من β_1، β_2 ؟

وإجابة هذا السؤال سوف توضح في الفصل الثالث.

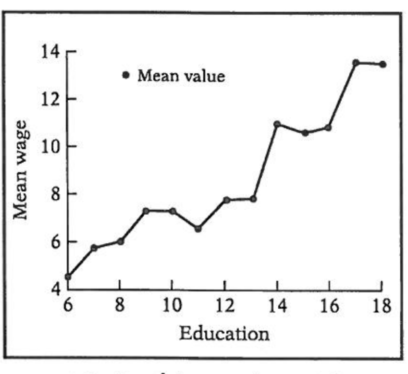
7.2 مثال توضيحي : AN ILLUSTRATIVE EXAMPLE

وسوف نلخص ماورد في هذا الفصل من خلال المثال التوضيحي التالي. جدول (6.2) يوضح المستوى التعليمي للفرد (مقاس بعدد السنوات الدراسية)،

ومتوسط أجر الساعة للفرد وفقاً لكل مستوى تعليمي، كذلك عدد الأفراد المناظر لكل مستوى تعليمي. وبيانات جدول (2–6) هي جزء قدمه Ernst Berndt من المسح السكاني population survey في مايو 1985[14]. وفي الفصل الثالث، سوف نتناول هذه البيانات بالتفصيل في المثال 3.3. وبرسم متوسط أجر الساعة مقابل مستوى التعليم، نحن نحصل على شكل (6.2). ومنحنى الانحدار في الشكل يوضح كيفية اختلاف متوسط الأجر وفقاً لمستوى التعليم. وفي الفصل التالي. سوف نوضح أنه توجد متغيرات أخرى غير المستوى التعليمي تؤثر على متوسط أجر الساعة.

جدول (6.2)

Years of schooling	Mean wage, $	Number of people
6	4.4567	3
7	5.7700	5
8	5.9787	15
9	7.3317	12
10	7.3182	17
11	6.5844	27
12	7.8182	218
13	7.8351	37
14	11.0223	56
15	10.6738	13
16	10.8361	70
17	13.6150	24
18	13.5310	31
	Total	528



شكل (6.2) متوسط أجر الساعة

8.2 الملخص والاستنتاجات : SUMMARY AND CONCLUSIONS

1 – يعتبر المفهوم الرئيسي لتحليل الانحدار هو دالة التوقع الشرطية expectation function (CEF) أو دالة انحدار المجتمع population regression function (PRF). وهدفنا من تحليل الانحدار هو كيفية الحصول على القيمة المتوسطة average value للمتغير التابع التي تتغير مع القيمة المعطاة للمتغير المفسر.

2 – هذا الكتاب يتناول بإسهاب دوال انحدار المجتمع الخطية linear PRE's، والخطية هنا تعني خطية في المعلمات parameters. وهذه الدوال ممكن أن تكون خطية أو غير خطية في المتغيرات المفسرة.

(14) Ernst R. Berndt, The Practice of Econometrics: Classic and Contemporary, Addison Wesley, Reading, Mass., 1991. Incidentally, this is an excellent book that the reader may want to read to find out how econometricians go about doing research.

3 - الحد العشوائي (أو المتغير العشوائي) u_i يلعب دوراً مهماً جداً في تقدير PRF.

4 - وبما أن الهدف الأساسي تقدير PRF باستخدام SRF، وبالتالي يصبح من الأهمية كيفية تقدير PRF عن طريق SRF، وهو ماسوف يتم تناوله في الفصل الثالث.

تمــــاريــن : EXERCISES

أسئلة : Questions

1.2 ماهي دالة التوقع الشرطي أو دالة انحدار المجتمع؟

2.2 ماهو الفرق بين دالة انحدار المجتمع ودالة انحدار العينة؟

3.2 ماهو دور حد الخطأ العشوائي stochastic error term u_i في تحليل الانحدار؟ وما هو الفرق بين حد الخطأ العشوائي u_i، والباقي the residual $\hat{u}_i$؟

4.2 لماذا نحتاج إلى تحليل الانحدار؟ ولماذا غير بسيط استخدام القيمة المتوسطة للمتغير التابع كأفضل قيمة؟

5.2 ماذا نعني بنموذج الانحدار الخطي؟

6.2 لكل نموذج من النماذج التالية، حدد أيًا منها يمثل نموذجاً خطياً في المعلمات، وأيها نموذج خطي في المتغيرات، وأيها نموذج خطي في المعلمات والمتغيرات؟ وماهي نماذج الانحدار الخطية؟

النموذج	العنوان الوصفي
a. $Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$	Reciprocal عكس
b. $Y_i = \beta_1 + \beta_2 \ln X_i + u_i$	Semilogarithmic شبه لوغاريتمي
c. $\ln Y_i = \beta_1 + \beta_2 X_i + u_i$	Inverse semilogarithmic شبه لوغاريتمي عكسي
d. $\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$	Logarithmic or double logarithmic لوغاريتمي- أولوغاريتمي
e. $\ln Y_i = \beta_1 - \beta_2 \left(\frac{1}{X_i}\right) + u_i$	Logarithmic reciprocal لوغاريتمي عكسي

ملحوظة: اللوغاريتم الطبيعي= ln (أي log =in)، u_i تعني المتغير العشوائي، وسوف نتناول هذه النماذج بالتفصيل في الفصل السادس.

7.2 هل النماذج التالية تعتبر نماذج انحدار خطي؟ ولماذا نعم ولماذا لا؟

a. $Y_i = e^{\beta_1 + \beta_2 X_i + u_i}$

b. $Y_i = \dfrac{1}{1 + e^{\beta_1 + \beta_2 X_i + u_i}}$

c. $\ln Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$

d. $Y_i = \beta_1 + (0.75 - \beta_1)e^{-\beta_2(X_i-2)} + u_i$

e. $Y_i = \beta_1 + \beta_2^3 X_i + u_i$

8.2 مـاذا نعني بنمـوذج الانحـدار الخطي؟ وإذا كـان $\beta_2 = 0.8$ في تمـرين d7.2، وهل النموذج خطي أو غير خطي؟

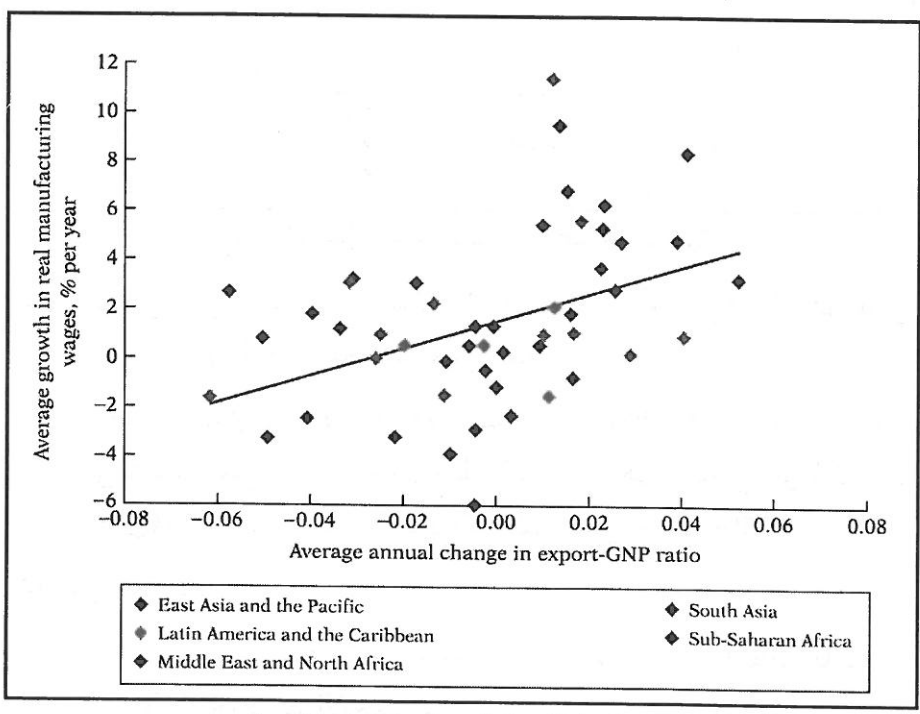
9.2 اعتبر النماذج التالية غير العشوائية (بمعنى عدم وجود المتغير العشوائي). هل تمثل هذه النماذج نماذج انحدار (بمعنى نماذج بدون المتغير العشوائي)؟

a. $Y_i = \dfrac{1}{\beta_1 + \beta_2 X_i}$

b. $Y_i = \dfrac{X_i}{\beta_1 + \beta_2 X_i}$

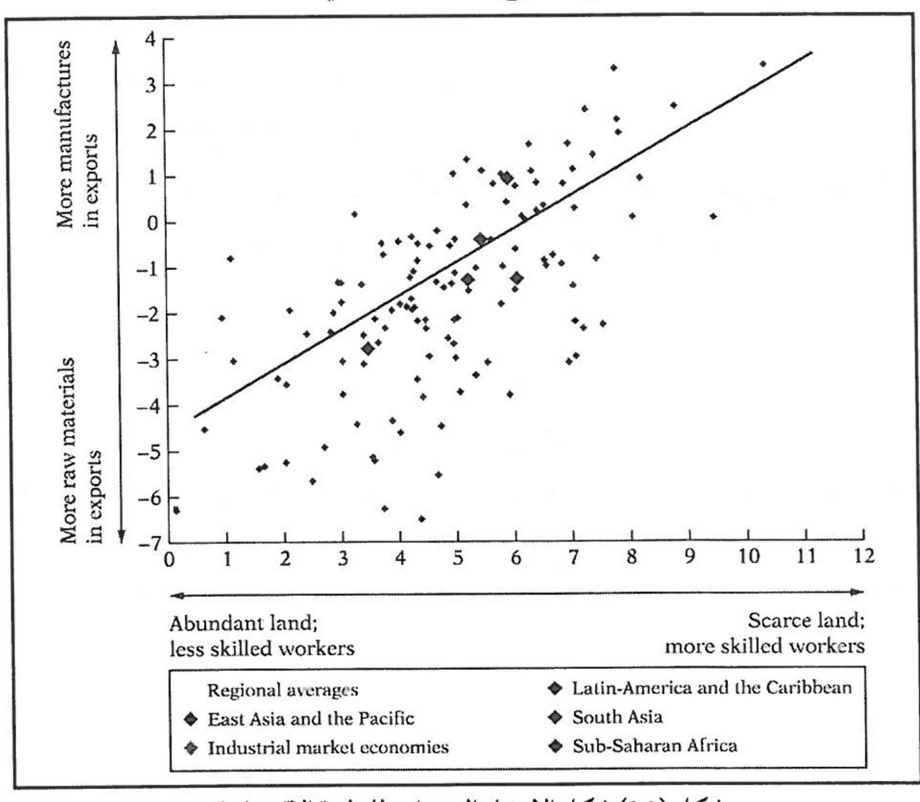
c. $Y_i = \dfrac{1}{1 + \exp(-\beta_1 - \beta_2 X_i)}$

10.2 اعـتبـر شكل الانتـشـار في الشكل التـالي مـاذا تسـتنتج من الشـكل؟ وهل خط الانحدار في الشكل يمثل خط انحدار مجتمع أو خط انحدار عينة؟



شكل (7.2) الانتشار العشوائي لنموذج الانحدار

11.2 اعتبر شكل الانتشار التالي ماذا تستنج من الشكل؟ ما النظرية الاقتصادية التي يمثلها الشكل؟ (انظر في مراجع الاقتصاد الدولي).



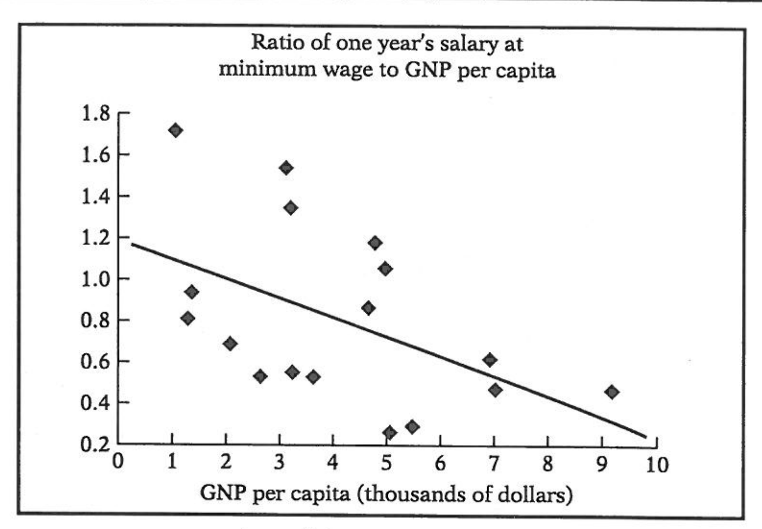
شكل (8.2) شكل الانتشار العشوائي للنظرية الاقتصادية

12.2 ماذا يمثل شكل الانتشار في شكل (9.2). وعلى أي أساس في الشكل يمكن إثبات أن قوانين الأجور المنخفضة تعتبر جيدة؟

13.2 هل خط الانحدار في شكل (3.1) في المقدمة يمثل PRF أم SRF؟ ولماذا؟ وكيف يمكن تفسير شكل الانتشار حول خط الانحدار، ماهي العوامل (أو المتغيرات الأخرى) المؤثرة على الإنفاق الاستهلاكي الشخصي بجانب GDP؟

14.2 اعتبر البيانات في جدول (7.2)

(أ) ارسم معدل مساهمة قوة العمل للذكور مقابل معدل البطالة للذكور. وبالنظر حدد خط الانحدار للشكل.



شكل (9.2) الانتشار بالنسبة للأجور المنخفضة

جدول (7.2)

LABOR FORCE PARTICIPATION DATA

Year	CLFPRM[1]	CLFPRF[2]	UNRM[3]	UNRF[4]	AHE82[5]	AHE[6]
1980	77.4	51.5	6.9	7.4	7.78	6.66
1981	77.0	52.1	7.4	7.9	7.69	7.25
1982	76.6	52.6	9.9	9.4	7.68	7.68
1983	76.4	53.9	9.9	9.2	7.79	8.02
1984	76.4	53.6	7.4	7.6	7.80	8.32
1985	76.3	54.5	7.0	7.4	7.77	8.57
1986	76.3	55.3	6.9	7.1	7.81	8.76
1987	76.2	56.0	6.2	6.2	7.73	8.98
1988	76.2	56.6	5.5	5.6	7.69	9.28
1989	76.4	57.4	5.2	5.4	7.64	9.66
1990	76.4	57.5	5.7	5.5	7.52	10.01
1991	75.8	57.4	7.2	6.4	7.45	10.32
1992	75.8	57.8	7.9	7.0	7.41	10.57
1993	75.4	57.9	7.2	6.6	7.39	10.83
1994	75.1	58.8	6.2	6.0	7.40	11.12
1995	75.0	58.9	5.6	5.6	7.40	11.44
1996	74.9	59.3	5.4	5.4	7.43	11.82

Source: Economic Report of the President, 1997. Table citations below refer to the source document.
[1]*CLFPRM,* Civilian labor force participation rate, male (%), Table B-37, p. 343.
[2]*CLFPRF,* Civilian labor force participation rate, female (%), Table B-37, p. 343.
[3]*UNRM,* Civilian unemployment rate, male (%) Table B-40, p. 346.
[4]*UNRF,* Civilian unemployment rate, female (%) Table B-40, p. 346.
[5]*AHE82,* Average hourly earnings (1982 dollars), Table B-45, p. 352.
[6]*AHE,* Average hourly earnings (current dollars), Table B-45, p. 352.

(ب) ماهي العلاقة المتوقعة بين المتغيرين، وما النظرية الاقتصادية وراء ذلك؟
وهل شكل الانتشار يتفق مع النظرية؟

(جـ) ارسم معدلات مساهمة عمل كل من الإناث والذكور مقابل متوسط مكسب الساعة (في 1982 بالدولار). (ممكن تستخدم كل شكل انتشار على حدة).

(د) هل يمكن رسم معدل مساهمة قوة العمل مقابل معدل البطالة ومتوسط مكسب الساعة معاً؟ كيف تصور العلاقة بين المتغيرات الثلاثة؟

15.2 جدول (8.2) يوضح الإنفاق على الطعام والإنفاق الكلي مقاس بالروبية، بالنسبة لعينة مكونة من 55 أسرة ريفية من الهند. (في بدايات 2000، الدولار الأمريكي يساوي 40 روبية هندية).

جدول (2-8) يمثل الإنفاق العام والكلي على الطعام

FOOD AND TOTAL EXPENDITURE (RUPEES)

Observation	Food expenditure	Total expenditure	Observation	Food expenditure	Total expenditure
1	217.0000	382.0000	29	390.0000	655.0000
2	196.0000	388.0000	30	385.0000	662.0000
3	303.0000	391.0000	31	470.0000	663.0000
4	270.0000	415.0000	32	322.0000	677.0000
5	325.0000	456.0000	33	540.0000	680.0000
6	260.0000	460.0000	34	433.0000	690.0000
7	300.0000	472.0000	35	295.0000	695.0000
8	325.0000	478.0000	36	340.0000	695.0000
9	336.0000	494.0000	37	500.0000	695.0000
10	345.0000	516.0000	38	450.0000	720.0000
11	325.0000	525.0000	39	415.0000	721.0000
12	362.0000	554.0000	40	540.0000	730.0000
13	315.0000	575.0000	41	360.0000	731.0000
14	355.0000	579.0000	42	450.0000	733.0000
15	325.0000	585.0000	43	395.0000	745.0000
16	370.0000	586.0000	44	430.0000	751.0000
17	390.0000	590.0000	45	332.0000	752.0000
18	420.0000	608.0000	46	397.0000	752.0000
19	410.0000	610.0000	47	446.0000	769.0000
20	383.0000	616.0000	48	480.0000	773.0000
21	315.0000	618.0000	49	352.0000	773.0000
22	267.0000	623.0000	50	410.0000	775.0000
23	420.0000	627.0000	51	380.0000	785.0000
24	300.0000	630.0000	52	610.0000	788.0000
25	410.0000	635.0000	53	530.0000	790.0000
26	220.0000	640.0000	54	360.0000	795.0000
27	403.0000	648.0000	55	305.0000	801.0000
28	350.0000	650.0000			

Source: Chandan Mukherjee, Howard White, and Marc Wuyts, *Econometrics and Data Analysis for Developing Countries*, Routledge, New York, 1998, p. 457.

(أ) ارسم البيانات على اعتبار أن المحور الرأسي يمثل الإنفاق على الطعام، والمحور الأفقي يمثل الإنفاق الكلي. وارسم خط الانحدار الذي يتوسط شكل الانتشار.

(ب) ماهي الاستنتاجات التي يمكن تحديدها من هذا المثال.

(جـ) هل تتوقع زيادة الإنفاق على الطعام زيادة خطية كلما زاد الإنفاق الكلي بدون النظر إلى مستوى الإنفاق الكلي؟ لماذا نعم ولماذا لا؟ يمكن استخدام الإنفاق الكلي بدلاً من الدخل الكلي.

16.2 جدول (9.2) يوضح درجات اختبار الموهبة الكلامية (SAT) لمجموعة من الأساتذة بإحدى الكليات خلال الفترة 1967–1996.

(أ) استخدم المحور الأفقي للسنويات، والمحور الرأسي لدرجات الاختبار لكل من الذكور والإناث كل على حدة.

(ب) ما الاستنتاجات العامة من الرسم.

(جـ) بمعرفة المستوى اللفظي لكل من الذكور والإناث، كيف يمكن التنبؤ بمستواهم في الرياضيات math.

(د) ارسم مستوى درجات الإناث اللفظي SAT، مقابل مستوى درجات الذكوراللفظي. ارسم خط الانحدار الذي يتوسط شكل الانتشار، وماهي ملاحظاتك؟

جدول (9.2) درجات اختبار الموهبة الكلامية (SAT)

MEAN SCHOLASTIC APTITUDE TEST SCORES FOR COLLEGE-BOUND SENIORS, 1967–1990*

	Verbal			Math		
Year	Males	Females	Total	Males	Females	Total
1967	463	468	466	514	467	492
1968	464	466	466	512	470	492
1969	459	466	463	513	470	493
1970	459	461	460	509	465	488
1971	454	457	455	507	466	488
1972	454	452	453	505	461	484
1973	446	443	445	502	460	481
1974	447	442	444	501	459	480
1975	437	431	434	495	449	472
1976	433	430	431	497	446	472
1977	431	427	429	497	445	470
1978	433	425	429	494	444	468
1979	431	423	427	493	443	467
1980	428	420	424	491	443	466
1981	430	418	424	492	443	466
1982	431	421	426	493	443	467
1983	430	420	425	493	445	468
1984	433	420	426	495	449	471
1985	437	425	431	499	452	475
1986	437	426	431	501	451	475
1987	435	425	430	500	453	476
1988	435	422	428	498	455	476
1989	434	421	427	500	454	476
1990	429	419	424	499	455	476

*Data for 1967–1971 are estimates.
Source: The College Board. The *New York Times*, Aug. 28, 1990, p. B-5.

الفصل الثالث

نموذج انحدار متغيرين.. مشكلة التقدير

TWO-VARIABLE REGRESSION MODEL: THE PROBLEM OF ESTIMATION

في الفصل الثاني، وجدنا أن مهمتنا الأولى هي تقدير دالة انحدار المجتمع (PRF) على أساس دالة انحدار العينة (SRF). وفي ملحق A، نحن نناقش بصورة عامة طريقتين للتقدير هما:

1 - طريقة المربعات الصغرى العادية: ordinary least squares (OLS).

2 - طريقة الإمكان الأعظم: maximum likelihood (ML).

وتعتبر طريقة المربعات الصغرى (OLS)، هي الطريقة الأكثر استخداماً في تحليل الانحدار. ومما هو جدير بالذكر أن استخدام طريقة المربعات الصغرى أو طريقة الإمكان الأعظم تعطيان نفس النتائج بالنسبة للانحدار الخطي.

1.3 طريقة المربعات الصغرى العادية: THE METHOD OF ORDINARY LEAST SQUARES

ترجع طريقة المربعات الصغرى العادية إلى عالم الرياضيات Carl Friedrich Gauss. وتحت فروض معينة (سوف نناقشها في الفصل 2.3)، تعتبر طريقة المربعات الصغرى أفضل طرق التقدير في تحليل الانحدار. ولتوضيح طريقة المربعات الصغرى، سوف نشرح أولاً أسس المربعات الصغرى:

بالرجوع إلى دالة انحدار المجتمع في متغيرين (PRF):

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.4.2)$$

وكما ذكرنا في الفصل الثاني أن الدالة PRF يتم تقديرها من الدالة SRF، ولم يتم إيجادها من بيانات المجتمع حيث:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \quad (2.6.2)$$

$$= \hat{Y}_i + \hat{u}_i \quad (2.6.3)$$

حيث $\hat{Y}_i$ هو توقع (التوقع الشرطي) لقيمة Y_i.

ولكن كيف يمكن تحديد الدالة SRF؟ وهو ما سوف يتم إنجازه فيما يلي. من الدالة (2.6.3) نجد أن:

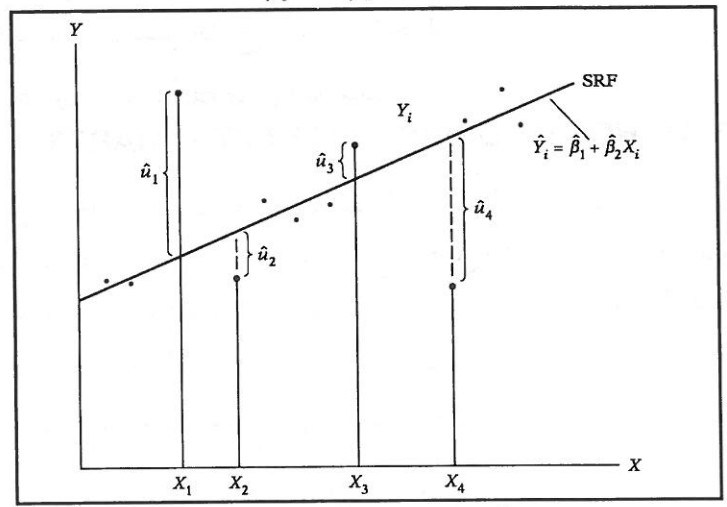
$$\hat{u}_i = Y_i - \hat{Y}_i$$

$$= Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i \quad (1.1.3)$$

والمعادلة (1.1.3) توضح أن البواقي $\hat{u}_i$ هي عبارة عن الفرق بين القيمة الفعلية والقيمة التقديرية لـ Y.

والآن إذا اعتبرنا عدد n من أزواج المشاهدات في المتغيرات X, Y أي وجود $(X_1,Y_1), (X_2,Y_2), ..., (X_n, Y_n)$ ويكون المطلوب هو إيجاد SRF باستخدام المشاهدات المعطاة بحيث تكون القيم المقدرة Y من SRF أي القيم $\hat{Y}_i$ أقرب مايمكن من القيم الفعلية لـ Y_i، ولإيجاد SRF التي تحقق الشرط السابق فيتحقق هذا الشرط عند المقدار $\Sigma\hat{u}_i$ حيث:

$$\sum_{i=1}^{n} \hat{u}_i = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)$$



شكل (3–1) كيفية تحديد الدالة SRF

أقل ما يمكن. ويعتبر هذا المعيار جيداً جداً لإيجاد SRF، كما سوف نوضح ذلك في شكل (1.3). ومن الشكل يتضح أن $\hat{u}_i$ تأخذ قيماً موجبة وقيماً سالبة مما يؤدي أن المجموع $\Sigma\hat{u}_i$ لا يعكس الاختلافات الفعلية $\hat{u}_i$ عن الدالة SRF لذلك يعتبر المقياس $\Sigma\hat{u}_i^2$ أفضل، حيث إن تأثير الإشارات الموجبة أو السالبة لـ $\hat{u}_i$ يتلاشى حيث:

$$\begin{aligned}\sum \hat{u}_i^2 &= \sum (Y_i - \hat{Y}_i)^2 \\ &= \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2\end{aligned} \qquad (2.1.3)$$

ومما هو جدير بالذكر، أن الحصول على القيم التقديرية $\hat{\beta}_1$ و $\hat{\beta}_2$ باستخدام طريقة المربعات الصغرى، أي باستخدام الطريقة التي تحقق النهاية الصغرى لـ $\sum_{i=1}^{n}\hat{u}_i^2$ تمكنا من الحصول على خصائص إحصائية مهمة، كما سوف نوضح فيما يلي:

من العلاقة (2.1.3) نجد أن:

$$\sum \hat{u}_i^2 = f(\hat{\beta}_1, \hat{\beta}_2) \qquad (3.1.3)$$

جدول (1.3) التحديد التجريبي للدالة SRF

Y_i (1)	X_i (2)	$\hat{Y}_{1i}$ (3)	$\hat{u}_{1i}$ (4)	$\hat{u}_{1i}^2$ (5)	$\hat{Y}_{2i}$ (6)	$\hat{u}_{2i}$ (7)	$\hat{u}_{2i}^2$ (8)
4	1	2.929	1.071	1.147	4	0	0
5	4	7.000	−2.000	4.000	7	−2	4
7	5	8.357	−1.357	1.841	8	−1	1
12	6	9.714	2.286	5.226	9	3	9
Sum: 28	16		0.0	12.214		0	14

Notes: $\hat{Y}_{1i} = 1.572 + 1.357X_i$ (i.e., $\hat{\beta}_1 = 1.572$ and $\hat{\beta}_2 = 1.357$)
$\hat{Y}_{2i} = 3.0 + 1.0X_i$ (i.e., $\hat{\beta}_1 = 3$ and $\hat{\beta}_2 = 1.0$)
$\hat{u}_{1i} = (Y_i - \hat{Y}_{1i})$
$\hat{u}_{2i} = (Y_i - \hat{Y}_{2i})$

إذا اعتبرنا البيانات X, Y في جدول (1–3)، وتم تقدير β_1 ، β_2 بطريقتين. الطريقة الأولى بحيث:

$$\hat{Y}_{1i} = 1.572 + 1.357X_1$$

أي $(\hat{\beta}_1 = 1.572, \hat{\beta}_2 = 1.357)$ ، والطريقة الثانية بحيث:

$$\hat{Y}_{2i} = 3.0 + 1.0X_i, \left(\hat{\beta}_1 = 3, \hat{\beta}_2 = 1.0\right)$$

ويتضح من الجدول أن مجموع مربعات البواقي في الحالة الأولى:

$$\sum \hat{u}_{1i}^2 = 12.214$$

في حين أن مجموع مربعات البواقي في الحالة الثانية:

$$\sum \hat{u}_{2i}^2 = 14$$

وهذا يعني أن التقدير بالطريقة الأولى أفضل من الطريقة الثانية، لأن مجموع مربعات البواقي أقل في الحالة الأولى.

ومن أهم خصائص طريقة المربعات الصغرى تحت فروض معينة، أنها الطريقة التي تعطينا أقل مجموع مربعات للبواقي، بالمقارنة بأي طريقة أخرى. لذلك تعتبر أفضل طريقة لإيجاد SRF كتقدير لـ PRF. وبالرجوع إلى الفصل الأول بملحق A3 أي بالرجوع إلى 1A-3، يتبين أنه يمكن باستخدام طريقة المربعات الصغرى الحصول على تقدير β_1 و β_2 أي الحصول على $\hat{\beta}_1$ و $\hat{\beta}_2$ من المعادلتين التاليتين:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_i \tag{4.1.3}$$

$$\sum Y_i X_i = \hat{\beta}_1 \sum X_i + \hat{\beta}_2 \sum X_i^2 \tag{5.1.3}$$

وتسمى المعادلتان السابقتان بالمعادلات الطبيعية normal equations، حيث n حجم العينة. وبحل المعادلتين السابقتين معاً نحصل على $\hat{\beta}_1$ و $\hat{\beta}_2$ على النحو التالي:

$$\hat{\beta}_2 = \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - \left(\sum X_i\right)^2} = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2} \tag{6.1.3}$$

حيث $\bar{X}, \bar{Y}$ هما متوسط كل من X, Y على الترتيب كذلك:

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \tag{7.1.3}$$

ومما هو جدير بالذكر أن التقديرات $\hat{\beta}_1$ و $\hat{\beta}_2$ باستخدام طريقة المربعات الصغرى العادية تحت فروض معينة تتميز بخصائص إحصائية [4] statistical properties:

1 - تقديرات المربعات الصغرى OLS عبارة عن دوال في قيم مشاهدات العينة، لذلك يتم حسابها بأسلوب بسيط وسهل.

(4) Rusell Davidson and James G. MacKinnon, Estimation and Inference in Econometrics, Oxford University Press, New York, 1993, p. 3.

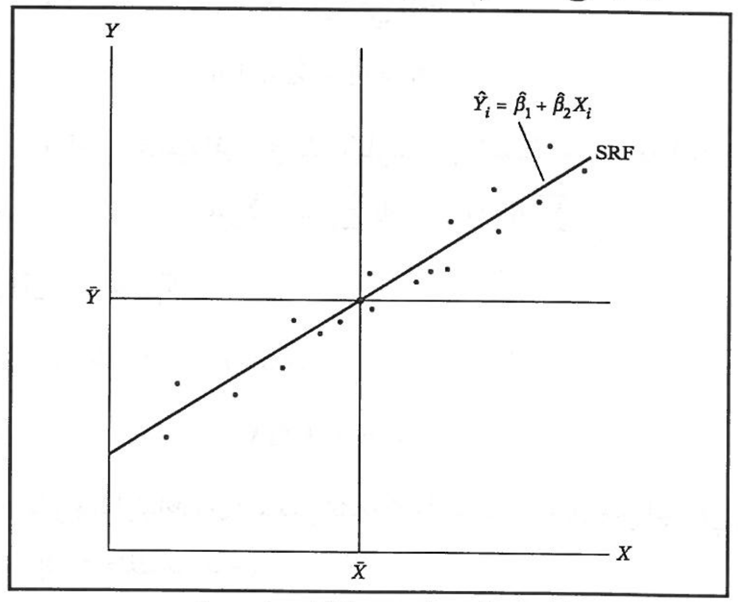
2 – تقديرات المربعات الصغرى تقديرات نقطة point estimators . من بيانات العينة (في الفصل الخامس، سوف نوضح نوعاً آخر من التقديرات يسمى بتقديرات الفترة interval estimators) .

3 – بما أن خط الانحدار يتم الحصول عليه من بيانات العينة، بالتالي فإن خط الانحدار يتميز بالخصائص التالية :

1 – يمر خط الانحدار بمتوسط كل من X, Y كما هو موضح في العلاقة (7.1.3) حيث :

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X}$$

كما هو موضح بيانياً في شكل (2.3)



شكل (2.3) خصائص خط الانحدار ومميزاته

2 – القيمة المتوسطة لـ $\hat{Y}$ (القيمة المتوسطة) تساوي $\hat{Y}_i$ القيمة المتوسطة لـ Y حيث :

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_i \\ &= (\bar{Y} - \hat{\beta}_2\bar{X}) + \hat{\beta}_2 X_i \qquad (9.1.3) \\ &= \bar{Y} + \hat{\beta}_2(X_i - \bar{X}) \end{aligned}$$

وبإجراء عملية المجموع على طرفي المعادلة السابقة، ثم القسمة على n، نحصل على المعادلة التالية:

$$\bar{\hat{Y}} = \bar{Y} \qquad [5](10.1.3)$$

3 – متوسط البواقي $\hat{u}_{1i}$ يساوي صفر. من ملحق A3 الفصل 3-1A نجد أن:

$$-2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2X_i) = 0$$

حيث $\hat{u}_i = Y_i - \hat{\beta}_i - \hat{\beta}_2X_i$ ، وبالتالي $-2\sum\hat{u}_i = 0$

وبالتالي: $\bar{\hat{u}} = 0$[6]

وكنتيجة لمعادلة انحدار العينة في (2.6.2):

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2X_i + \hat{u}_i \qquad (2.6.2)$$

وبإجراء عملية المجموع على الطرفين ثم القسمة على n نجد أن:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2\sum X_i + \sum\hat{u}_i \qquad (11.1.3)$$

وبما أن $\sum\hat{u}_i = 0$

وبقسمة الطرفين على n نجد أن:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2\bar{X} \qquad (12.1.3)$$

4 – لايوجد ارتباط بين المتغير التابع Y والمتغير $\hat{u}$ الذي يمثل البواقي $\hat{u}_i$ ويمكن إثبات ذلك من ملحق A3.

كذلك المتغير العشوائي $\hat{u}$ غير مرتبط بالمتغير المفسر X. ويمكن إثبات ذلك أيضاً من ملحق A3.

(5) Note that this result is true only when the regression model has the intercept term b1 in it. As App. 6A, Sec. 6A.1 shows, this result need not hold when b1 is absent from the model.

(6) This result also requires that the intercept term b1 be present in the model (see App. 6A, sec. 6A.1).

2.3 نموذج الانحدار الخطي التقليدي.. فروض طريقة المربعات الصغرى :

THE CLASSICAL LINEAR REGRESSION MODEL:

THE ASSUMPTIONS UNDERLYING THE METHOD OF LEAST SQUARES

إن هدفنا من تحليل الانحدار ليس فقط الحصول على $\hat{\beta}_1$ و$\hat{\beta}_2$ باستخدام طريقة المربعات الصغرى، ولكن هدفنا هو إجراء استدلال إحصائي حول كل من $\hat{\beta}_1$ و$\hat{\beta}_2$. فعلى سبيل المثال، كيف يمكن الحصول على $\hat{Y}_i$ التي تكون أقرب ما يمكن إلى $E(Y \mid X_i)$. ولمعرفة ذلك لابد من الإلمام بالفروض assumptions التي يجب توافرها للحصول على $\hat{Y}_i$ التي تكون أقرب مايمكن من Y_i . من دالة انحدار المجتمع PRF أو المتغير التابع Y_i في الدالة التالية :

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

يعتمد على كل من $X_i \mid u_i$. وبالتالي الاستدلال الإحصائي حول Y_i يجب أن يتناول الفروض عن المتغير (أو المتغيرات) X_i، والمتغير u_i.

ونموذج الانحدار الخطي التقليدي classical linear regression model (CLRM) أو مايسمى بنموذج الانحدار المعياري standard mode أو نموذج جاوس Gaussian model . ويعتبر هذا النموذج حجر الزاوية في نظرية الاقتصاد القياسي econometric theory عند توافر الـ10 فروض التالية[7] .

وفيما يلي، سوف نناقش هذه الفروض في إطار نموذج الانحدار في متغيرين (متغير تابع ومتغير واحد مفسر)، وفي الفصل السابع، سوف نتناول هذه الفروض بالنسبة لنموذج الانحدار المتعدد.

الفرض (1): نموذج الانحدار الخطي، يعتبر نموذج انحدار خطي في المعنمات parameters كما في العلاقة (2.4.2) على النحو:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (2.4.2)$$

(7) It is classical in the sense that it was developed first by Gauss in 1821 and since then has served as a norm or a standard against which may be compared the regression models that do not satisfy the Gaussian assumptions.

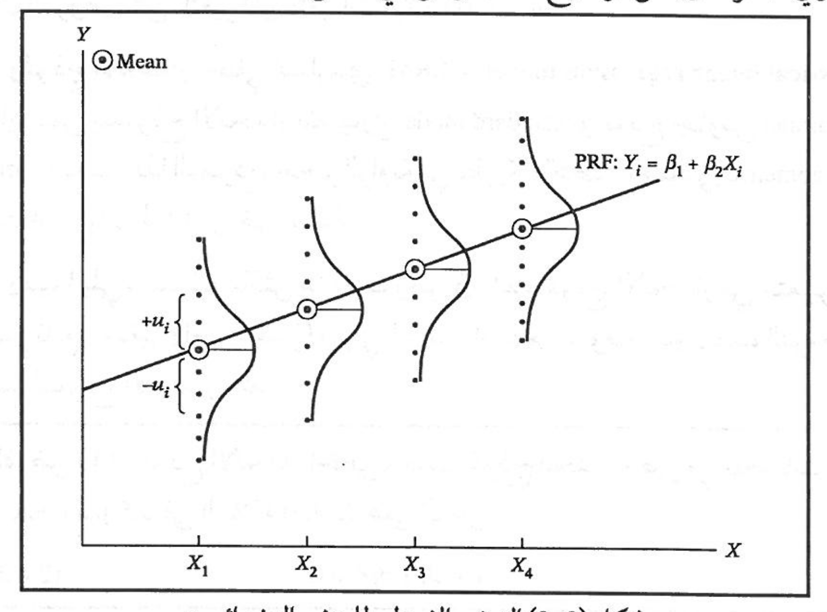
ومما هو جدير بالذكر، أن نموذج الانحدار الخطي في المعلمات β_1 ، β_2 إممكن أن يكون غير خطي في Y ، X كما سبق مناقشة ذلك في الفص الثاني[8].

الفرض (2): المتغير X يأخذ قيماً غير عشوائية nonstochastic أو قيماً محددة fixed في المعاينة التكرارية repeated sampling.

ويعتبر افتراض أن X يأخذ قيماً محددة فرض مهم في التحليل.

الفرض (3): القيمة المتوقعة (المتوسطة) mean value للمتغير العشوائي u_i تساوي صفراً عند أي قيمة من قيم X أو بعبارة أخرى:
$E(u_j \mid X_j) = 0$ (1.2.3)
وبالتالي $E(u_j) = 0$ [9]

والفرض (3) يحدد أن التوقع الشرطي لـ u_j عند أي مستوى من مستويات X_i تساوي صفراً. ويمكن توضيح هذا الفرض في شكل (3.3).



شكل (3-3) التوزيع الشرطي للمتغير العشوائي u_i

(8) However, a brief discussion of nonlinear-in-the-parameter regression models is given in Chap. 17.

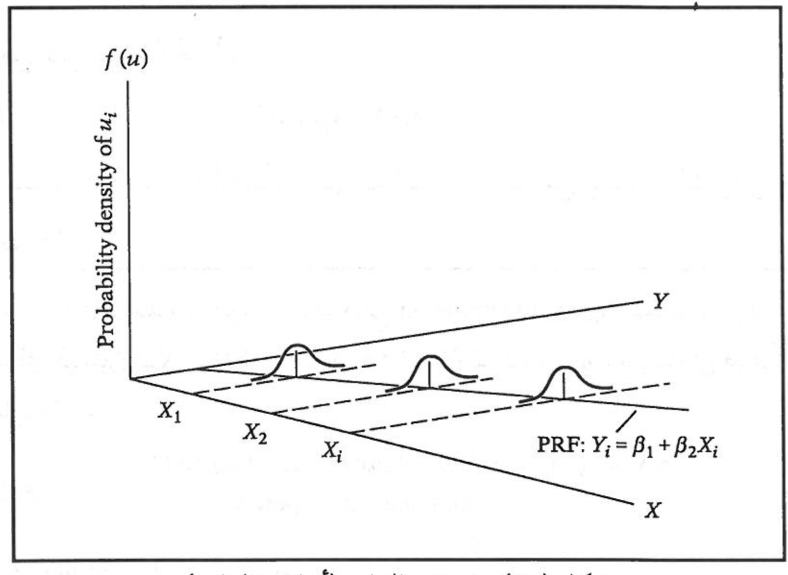
(9) For illustration, we are assuming merely that the u's are distributed symmetrically as shown in Figure 3.3. But shortly we will assume that the u's are distributed normally.

الفرض (4): تساوي التباين للمتغيرات u عند أي مستوى للمتغير X_i أو بعبارة أخرى.

$$\begin{aligned}\textbf{var}\,(u_i \mid X_i) &= E[u_i - E(u_i \mid X_i)]^2 \\ &= E\left(u_i^2 \mid X_i\right) \text{ because of Assumption 3} \qquad \textbf{(2.2.3)} \\ &= \sigma^2\end{aligned}$$

حيث var تشير إلى التباين

والمعادلة (2.2.3) تحدد أن تباين u_i عند كل مستوى X_i (بمعنى التباين الشرطي لـ u_i) وشكل (4.3) يوضح فرض تساوي التباينات للمتغير u_i عند أي مستوى من مستويات X_i.

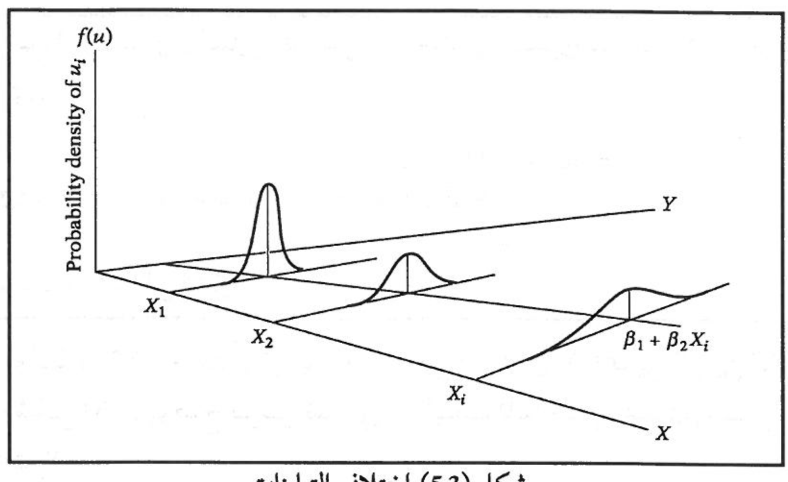


شكل (4.3) عدم تغير التباين (أو ثبات التباين)

كذلك شكل (5.3) يوضح اختلاف التباين للمتغير u_i عند المستويات المختلفة للمتغير X_i أي عندما:

$$\text{var}\,(u_i \mid X_i) = \sigma_i^2 \qquad \textbf{(3.2.3)}$$

(10) For a more technical reason why Assumption 3 is necessary see E. Malinvaud, Statistical Methods of Econometrics, Rand McNally, Chicago, 1966, p. 75. See also exercise 3.3.



شكل (5.3) اختلاف التباينات

ومن الفرض (4) نجد أن

$$\text{var}\,(Y_i \mid X_i) = \sigma^2 \qquad \textbf{(5.2.3)}$$

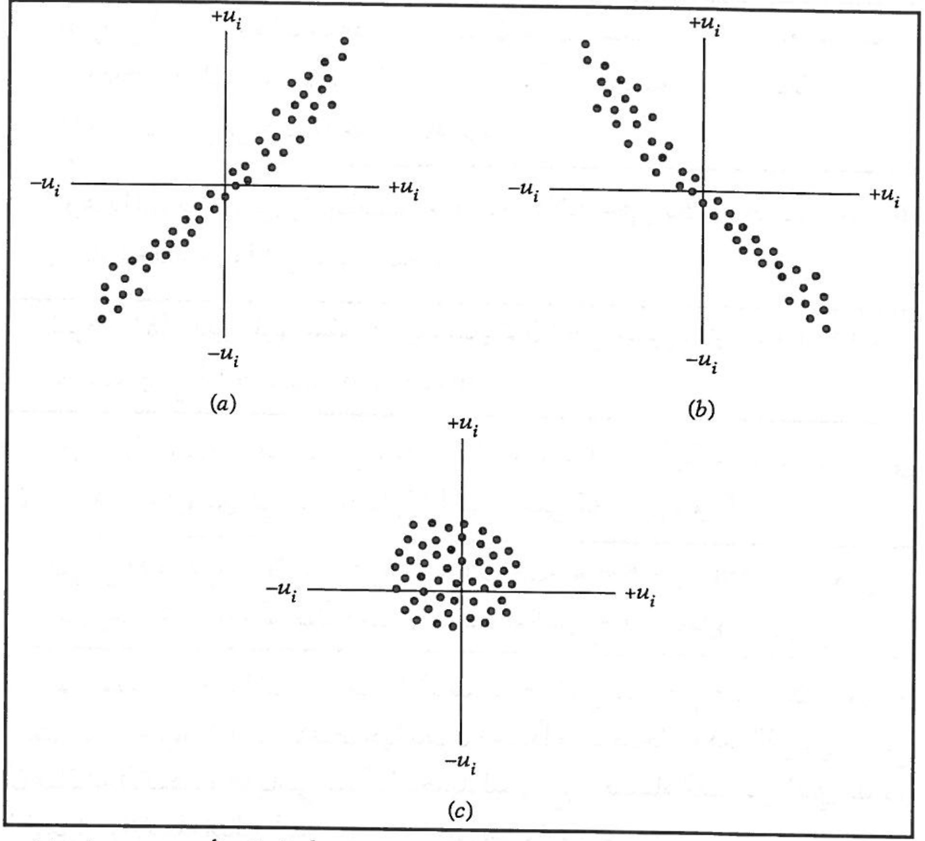
كذلك $\text{var}\,(Y) = \sigma^2$ أيضاً. يعطي تفاصيل أكثر عن التباين الشرطي وغير الشرطي لـ Y.

الفرض (5): عدم وجود ارتباط ذاتي autocorrelation بين المتغيرات u_i. أي عند أي قيمتين لـ X، ولتكن X_i، X_j، $i \neq j$، الارتباط بين u_i , u_j يساوي صفراً أو بعبارة أخرى

$$\begin{aligned} \textbf{cov}\,(u_i, u_j \mid X_i, X_j) &= E\{[u_i - E(u_i)] \mid X_i\}\{[u_j - E(u_j)] \mid X_j\} \\ &= E(u_i \mid X_i)(u_j \mid X_j) \qquad \text{(why?)} \qquad \textbf{(5.2.3)} \\ &= 0 \end{aligned}$$

حيث cov تشير إلى covariance

من العلاقة (5.2.3) يتضح أن المتغيرين u_i , u_j غير مرتبطين uncorrelated. وهذا الفرض لعدم وجود ارتباط ذاتي بين u_i , u_j. وشكل (6.3) يوضح الأشكال المختلفة في حالة وجود ارتباط ذاتي موجب في شكل (a.6.3) أو سالب في (b.6.3) أو عدم وجود ارتباط ذاتي في شكل (b.6.3).



شكل (6.3) يوضح النماذج المختلفة للارتباط الخطي الموجب أو السالب أو عدم وجود ارتباط

الفرض (6): التغير بين u_i , X_i يساوي صفراً، أو بعبارة أخرى $E(u_iX_j) = 0$ ،

حيث :

$$\begin{aligned} \text{cov}(u_i, X_i) &= E[u_i - E(u_i)][X_i - E(X_i)] \\ &= E[u_i(X_i - E(X_i))] \quad \text{since } E(u_i) = 0 \\ &= E(u_iX_i) - E(X_i)E(u_i) \quad \text{since } E(X_i) \text{ is nonstochastic} \\ &= E(u_iX_i) \quad \text{since } E(u_i) = 0 \\ &= 0 \quad \text{by assumption} \end{aligned} \tag{6.2.3}$$

وفرض (6) يؤكد على عدم وجود ارتباط بين المتغير المفسر X والمتغير العشوائي u. والفرض (6) يفترض أن المتغيرين X، u غير مرتبطين أو مستقلين، بمعنى أن تأثير كل منهما على المتغير التابع Y تأثير منفصل. ولكن في حالة إذا كان يوجد ارتباط بين X , u فإنه في هذه الحالة لايمكن أن يكون تأثير X على Y منفصلاً عن تأثير u على u، أو بعبارة أخرى لايمكن وضع Y كدالة جمعية (أي مجموع) في المتغيرين X، u.

الفـرض (7): عـدد المشـاهدات يجب أن يكون أكـبـر من عـدد المعلمـات parameters المطلوب تقـديرها. أو بعـبـارة أخـرى، يجب أن يكون عـدد المشاهدات أكبر من عدد المتغيرات المفسرة.

وهـذا الفرض ضـرورة ريـاضـيـة لحـل المعـادلات حتى يمكـن الحصـول على $\hat{\beta}_1$ و $\hat{\beta}_2$ باستخدام طريقة المربعات الصغرى.

الفرض (8): تغيير قيم المتغير X، بحيث يكون تباين المتغير X أي var (X) عدد موجب نهائي finite positive number[13].

في حالة عـدم توافـر فـرض (8)، وكـانت قـيم X_i متسـاوية، فـهـذا يـؤدي إلى أن $X_i = \bar{X}$ وهذا يؤدي إلى عدم إمكانية الحصول على كل من $\hat{\beta}_1$ و $\hat{\beta}_2$.

الفرض (9): نموذج الانحدار نموذج محدد وصحيح correctly specified. بمعنى عدم وجود تحيز bias أو خطأ error في استخدام النموذج في التحليل التجريبي.

في المقدمة ناقشنا أن منهاجية الاقتصاد القياسي تفترض ضمناً أن النموذج المستخدم لاختبار النظرية الاقتصادية يكون محدداً وصحيحاً. وهذا الفرض يوضح أن مناقشة الاقتصاد القياسي تبدأ بالتحديد لنموذج الاقتصاد القياسي الذي يصور الظاهرة. **والأسئلة التالية تحدد النموذج الذي يتم بناؤه وهي:**

1 - ماهي المتغيرات التي يتضمنها النموذج؟

2 - ماهي الصياغة الرياضية للنموذج؟

3 - ماهي الفروض الاحتمالية حول كل من Y_i، X_i | u_i؟

وهذه الأسئلة الثلاثة، سوف تناقش بالتفصيل في الفصل (13)، وإلغاء متغيرات مهمة في النموذج أو اختيار صياغة رياضية غير مناسبة، أو فروض احتمالية غير صحيحة. فكل هذا سوف يؤدي إلى تقديرات غير ملائمة. ويمكن توضيح ذلك من خلال شكل (1.3). بافتراض وجود صياغتين للعلاقة بين معدل تغير الأجر النقدي Y_i ومعدل البطالة (Y) على النحو التالي:

---

(13) The sample variance of X is

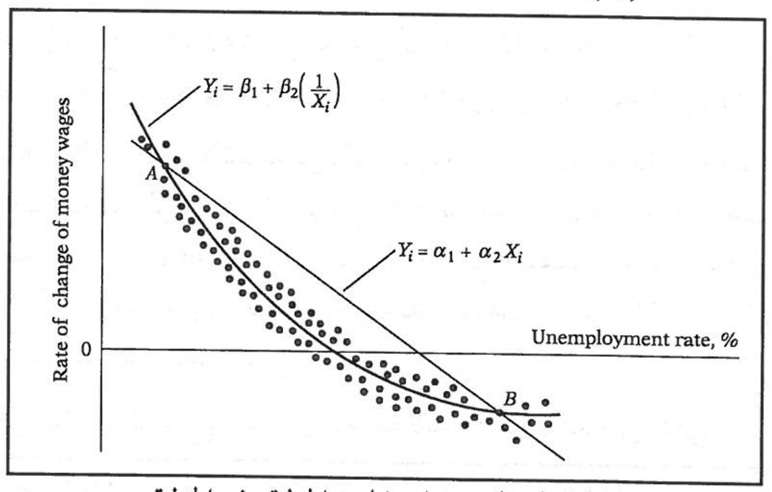
$$\text{var}(X) = \frac{\sum(X_i - \bar{X})^2}{n-1}$$

where n is sample size.

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \tag{7.2.3}$$

$$Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i \tag{8.2.3}$$

وشكل (7.3) يوضح شكل الانتشار ، كذلك شكل النموذجين لكل من (7.2.3)، (8.2.3). من الشكل يتضح أن استخدام النموذج $\hat{Y} = \hat{\alpha}_1 + \hat{\alpha}_2 X_i$. في التقدير يعتبر تحديداً خاطئاً Specification error ، وتحديداً متحيزاً specification bias . حيث إن النموذج $\hat{Y}_i = \hat{\beta}_2 + \hat{\beta}_1\left(\frac{1}{X_i}\right)$ يمثل البيانات أفضل كما هو واضح في الشكل .



شكل (3-7) منحنيات فيلبس الخطية وغير الخطية

ولتحديد المتغيرات التي يتضمنها النموذج ، وشكل العلاقة الرياضية للنموذج ، يجب إجراء تحليل تجريبي empirical analysis(14) .

مما سبق ، يتضح أن صحة التحليل ونتائجه مرتبطة تماماً بصحة الإجابة على الأسئلة السابقة ، وذلك بصفة خاصة عندما توجد نظريات متعددة تحاول شرح الظاهرة الاقتصادية economic phenomenon . وعلى سبيل المثال لذلك ، معدل التضخم inflation rate ، أو الطلب على النقود the demand for money ، أو القيمة

(14) But one should avoid what is known as **"data mining,"** that is, trying every possible model with the hope that at least one will fit the data well. That is why it is essential that there be some economic reasoning underlying the chosen model and that any modifications in the model should have some economic justification. A purely ad hoc model may be difficult to justify on theoretical or a priori grounds. In short, theory should be the basis of estimation. But we will have more to say about data mining in Chap. 13, for there are some who argue that in some situations data mining can serve a useful purpose.

التوازنية للسهم والسند .equilibrium value of a stock or a bond مما سبق يتضح أن بناء نموذج الاقتصاد القياسي غالباً هو فن أكثر منه علم.

ومما سبق، يتضح أنه من الأهمية أن نلاحظ أن جميع الفروض السابقة تتعلق بدالة انحدار المجتمع PRF وليس دالة انحدار العينة SRF. ومن الأهمية أن نلاحظ أن طريقة المربعات الصغرى السابق مناقشتها لها نفس الخصائص التي تتشابه مع الفروض الموضوعة على PRF. فعلى سبيل المثال، أن إيجاد $\Sigma u_i = 0$، كذلك $\bar{u} = 0$ تتشابه مع افتراض أن $E(u_i \mid X_i) = 0$.

بالمثل إيجاد $\Sigma \hat{u}_i X_i = 0$ مشابه لافتراض أن $\text{cov}\,(u_i \mid X_i) = 0$. وبالتالي فإنا نلاحظ أن طريقة المربعات الصغرى تحاول أن تتشابه مع الفروض المفروضة علىPRF.

وبالطبع الدالة SRF لاتشابه في ذلك الفروض للنموذج CLRM. وكما سوف نوضح فيما بعد، بالرغم من أن افتراض $\text{cov}\,(u_i \mid u_j) = 0$، $i \neq j$، أنه غير حقيقي أن في العينة $\text{cov}\,(\hat{u}_i \mid \hat{u}_j) = 0$، $i \neq j$. ولكن المشكلة في الحقيقة، نحن سوف نوضح مؤخراً، أن البواقي u_i ليس فقط مرتبطة ذاتياً autocorrelated ولكن ذات تباينات مختلفة وليست ثابتة (انظر الفصل 12).

وفي حالة تناول نماذج الانحدار المتعددة لابد من تناول الفرض التالي.

الفرض (10): عدم وجود تداخل خطي تام perfect multicollinearity. أي عدم وجود علاقة خطية تامة بين المتغيرات المفسرة X. وسوف يناقش هذا الفرض بالتفصيل في الفصل السابع، حيث يتم تناول نماذج الانحدار المتعددة.

كلمة عن هذه الفروض: A Word about These Assumptions

وسؤال يطرح نفسه على جانب بالغ من الأهمية وهو: ماهي واقعية هذه الفروض؟ بمعنى هل تتفق هذه الفروض مع واقع المشاكل الفعلية أم لا؟ كذلك في كثير من الحالات لاتتوافر بعض أو كل هذه الفروض[15]. ومما هو جدير بالذكر، أن وضع فروض معينة certain assumptions وذلك بهدف تطوير المسألة الموضوعية Subject matter في خطوات تدريجية gradual steps وليس لأن هذه الفروض واقعية[16].

(15) Milton Friedman, Essays in Positive Economics, University of Chicago Press, Chicago, 1953, p. 14.

(16) Mark Blaug, The Methodology of Eeonomirs: Or How Economists Explain, 2d ed., Cambridge University Press, New York, 1992, p. 92.

وفي الفصول التالية ، سوف نوضح خصائص نموذج CLRM في حالة عدم وجود فرض أو أكثر من الفروض السابقة. وفي نهاية هذا الفصل ، سوف نوضح في جدول (3-4) إرشاداً لتحديد مايحدث للنموذج CLRM في حالة عدم تحقق فرض أو أكثر.

ونظراً لدراسة نموذج CLRM، فإنه يكون من الأهمية دراسة الخصائص الإحصائية statistical properties لطريقة المربعات الصغرى العادية OLS مقارنة بالخصائص العددية numerical properties السابق مناقشتها. علماً بأن الخصائص الإحصائية لطريقة المربعات الصغرى مبنية على فروض نموذج CLRM التي قدمها عالم الرياضيات جاوس ماركوف في نظريته Gauss- Markov theorem. وقبل تقديم نظرية ماركوف يعتبر من الأهمية تقديم مفهومي الدقة precision أو الأخطاء المعيارية standard errors بالنسبة لتقديرات المربعات الصغرى.

3.3 الدقة أو الأخطاء المعيارية لتقديرات المربعات الصغرى : PRECISION OR STANDARD ERRORS OF LEAST SQUARES ESTIMATES

من المعادلتين (6.1.3) ، (7.1.3) يتضح أن تقديرات المربعات الصغرى هي دوال في بيانات العينة. ونظراً لاختلاف البيانات من عينة لعينة أخرى، وبالتالي فإن التقديرات التي يتم الحصول عليها من بيانات عينة معينة تختلف عن التقديرات التي يتم الحصول عليها من عينة أخرى. وبالتالي فأي تقديرات يفضل استخدامها؟.

ولتحديد أفضل تقديرات، فإن ذلك لايتطلب الرجوع لمقياس الصلاحية reliability measure أو مقياس الدقة precision measure للمفاضلة بين التقديرات المختلفة $\hat{\beta}_1$ و $\hat{\beta}_2$ من العينات المختلفة. وفي الإحصاء القياسي دقة التقدير باستخدام الأخطاء المعيارية standard errors[17] للتقدير. وتحت فروض جاوس، كما هو موضح بملحق 3A3 في الفصل (3)، أن الأخطاء المعيارية لتقديرات المربعات

(17) The **standard error** is nothing but the standard deviation of the sampling distribution of the estimator, and the sampling distribution of an estimator is simply a probability or frequency distribution of the estimator, that is, a distribution of the set of values of the estimator obtained from all possible samples of the same size from a given population. Sampling distributions are used to draw inferences about the values of the population parameters on the basis of the values of the estimators calculated from one or more samples. (For details, see **App. A.**)

الصغرى يتم الحصول عليها باستخدام المعادلات التالية:

$$\mathrm{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad (1.3.3)$$

$$\mathrm{se}(\hat{\beta}_2) = \frac{\sigma}{\sqrt{\sum x_i^2}} \quad (2.3.3)$$

$$\mathrm{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \quad (3.3.3)$$

$$\mathrm{se}(\hat{\beta}_1) = \sqrt{\frac{\sum X_i^2}{n\sum x_i^2}}\sigma \quad (4.3.3)$$

حيث var تشير إلى التباين، se تشير إلى الخطأ المعياري، σ^2 تشير إلى تباين u_i كما هو وارد في الفرض (4).

هناك $x_i = X_i - \bar{X}$

يتم حسابها من بيانات العينة باستثناء σ^2، حيث يمكن تقديره من البيانات أيضاً على النحو التالي:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \quad (5.3.3)$$

كماهو موضح في ملحق 3A3 بالفصل (A3).

حيث $\hat{\sigma}^2$ تقدير المربعات الصغرى OLS للمقدار σ^2 حيث يعرف $(n-2)$ بعدد درجات الحرية number of degrees of freedom (df)، والمقدار $\sum \hat{u}_i^2$ يمثل مجموع مربعات البواقي residual sum of squares (RSS)[18].

وبمعرفة $\sum \hat{u}_i^2$ يمكن حساب $\hat{\sigma}^2$. $\sum \hat{u}_i^2$ يمكن حسابها من العلاقة (2.1.3) أو من العلاقة التالية:

(18) The term **number of degrees of freedom** means the total number of observations in the sample (= n) less the number of independent (linear) constraints or restrictions put on them. In other words, it is the number of independent observations out of a total of n observations. For example, before the RSS (3.1.2) can be computed, $\hat{\beta}_1$ and $\hat{\beta}_2$ must frst be obtamed. These two estimates therefore put two restrictions on the RSS. Therefore, there are n - 2, not n, independent observations to compute the RSS. Following this logic, in the three-variable regression RSS will have n - 3 df, amd for the k-variable model it will have n - k df. The general rule rs tbis: df = (n - number of parameters estimated).

$$\sum \hat{u}_i^2 = \sum (y_i - \bar{y})^2 - \hat{\beta}_2^2 \sum (x_i - \bar{X}) \tag{6.3.3}$$

وبمقارنة المعادلتين (6.3.3) و (2.1.3) يتضح بسهولة أنه ليس بالضرورة حساب $\hat{u}_i$ لكل مشاهدة، ولكن يمكن إيجاد $\hat{u}_i^2$ باستخدام الطرف الأيمن للمعادلة (6.3.3)، كما سوف نوضح ذلك في الفصلين 11 ، 12).

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

وبما أن التعبير البديل لحساب $\sum \hat{u}_i^2$ كما هو موضح :

$$\sum \hat{u}_i^2 = \sum y_i^2 - \frac{\left(\sum x_i y_i\right)^2}{\sum x_i^2} \tag{7.3.3}$$

حيث $y_i = Y_i - \bar{Y}$

وبالتالي فإن

$$\hat{\sigma} = \sqrt{\frac{\sum \hat{u}_i^2}{n-2}} \tag{8.3.3}$$

حيث ترمز $\hat{\sigma}$ إلى الخطأ المعياري للتقدير the standard error of estimate أو الخطأ المعياري للانحدار the standard error of the regression (se) . حيث يعتبر $\hat{\sigma}$ تقديراً للانحراف المعياري لقيم Y حول خط الانحدار المقدر. وغالباً يستخدم كمقياس لجودة التوفيق goodness of fit لخط انحدار النموذج.

وكما ذكرنا سابقاً أن X_2، σ^2 تمثل التباين الشرطي the (conditional) variance لكل من u_i، Y_i . لذلك الخطأ المعياري للتقدير يمكن أن يسمى أيضاً بالانحراف الشرطي لـ u_i أو Y_i . وبالطبع عادة، σ_Y^2, σ_Y تمثل على الترتيب التباين غير الشرطي، والانحراف المعياري غير الشرطي للمتغير Y.

وفيما يلي خصائص تباين (وكذلك الانحراف المعياري) لكل من $\hat{\beta}_1$ و $\hat{\beta}_2$.

1 - تباين $\hat{\beta}_2$ يتناسب مع σ^2 ولكن يتناسب عكسياً مع $\sum x_i^2$.

2 - تباين $\hat{\beta}_1$ يكون متناسبًا مع σ^2، $\sum x_i^2$ ولكن متناسب عكسياً مع $\sum x_i^2$ وحجم العينة n.

3 - وبما أن $\hat{\beta}_1$ و $\hat{\beta}_2$ تقديران، ومن نفس العينة، بالتالي فهما غير مستقلين، ويمكن قياس ذلك بمقياس التغير covariance بينهما والذي نرمز له بالرمز $\text{cov}(\hat{\beta}_2, \hat{\beta}_1)$ كما سوف نوضح في ملحق 4A3 بالفصل (3) حيث:

$$\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2) = -\bar{X}\,\mathrm{var}(\hat{\beta}_2) = -\bar{X}\left(\frac{\sigma^2}{\sum x_i^2}\right) \quad (9.3.3)$$

وبما أن تباين $\hat{\beta}_2$ أي $\mathrm{var}(\hat{\beta}_2)$ دائماً موجب، بالتالي فإن إشارة التغير بين $\hat{\beta}_1$ و $\hat{\beta}_2$ تعتمد على إشارة $\bar{X}$. فإذا كان $\bar{X}$ موجبًا فإن التغاير يكون سالبًا أي العلاقة بين $\hat{\beta}_1$ و $\hat{\beta}_2$ علاقة عكسية، وفيما بعد، سوف نوضح أهمية دراسة أهمية التغير بين معاملات الانحدار. ولكن كيف تعتبر التباينات variances والأخطاء المعيارية standard errors قادرة على توضيح صلاحية reliability هذه التقديرات؟ وهذه تعتبر مشكلة في الاستدلال الإحصائي statistical inference كما سوف نوضح في الفصلين (4، 5).

4.3 خصائص تقديرات المربعات الصغرى :

PROPERTIES OF LEAST SQUARES ESTIMATORS:

نظرية جاوس ماركوف : THE GAUSS- MARKOV THEOREM[19]

وكما سبق، فإن فروض نموذج الانحدار التقليدي، وتقديرات المربعات الصغرى، تعتبر تقديرات ذات خصائص مثلى optimum properties. هذه الخصائص تتضمنها نظرية جاوس ماركوف. ولفهم هذه النظرية، نحن نحتاج إلى خاصية أفضل خط غير متحيز best linear unbiasedness property للتقدير[20]. وكما أوضحنا في ملحق A. يقال للتقدير OLS تقدير $\hat{\beta}_2$ أنه تقدير أفضل خط غير متحيز (BLUE) للمعلمة $\hat{\beta}_2$ في حالة:

1 - يكون خط أي دالة خطية للمتغير العشوائي، بمعنى أن المتغير العشوائي التابع Y في نموذج الانحدار.

2 - يكون غير متحيز، بمعنى أن المتوسط أو توقع $\hat{\beta}_2$ بمعنى $\mathrm{E}(\hat{\beta}_2)$ يساوي المعلمة الحقيقية $\hat{\beta}_2$.

3 - للتقدير $\hat{\beta}_2$ أقل تباين minimum variance في فئة التقديرات الخطية غير المتحيزة.

وفي مفهوم الانحدار يمكن إثبات أن تقديرات OLS تكون BLUE. وهذا ما تقره نظرية جاوس ماركوف.

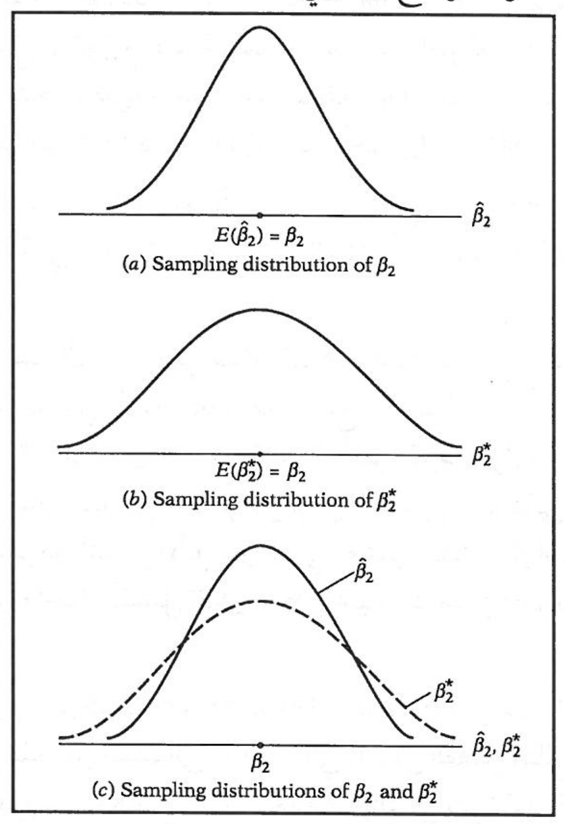
---

(19) Although known as the Gauss-Markov theorem, the least-squares approach of Gauss antedates (1821) the minimum-variance aproach of Markov (1900),

(20) The reader should refer to **App. A** for the importance of linear estimators as well as for a general discussion of the desirable properties of statistical estimators.

نظرية جـاوس مـاركـوف: تحت فـروض نموذج الانحـدار الخطي تقـديرات المربعات الصغرى، في فئة التقديرات الخطية غير المتحيزة لها أقل تباين، ويشار لهما بتقديرات BLUE.

وإثبات هذه النظرية موضح في ملحق 6A3، بالفـصل (3). والأهميـة الكامنة لنظرية ماركوف سوف توضح فيما يلي:



شكل (3-8) توزيع المعاينة للتقدير OLS $\hat{\beta}_2$ وتقديرات β_2^* البديلة

وأنه يكفي أن نلاحظ أن النظرية theorem لهـا مـحتـوى نظري theoretical ذو أهمية عملية practical importance[21]. ويمكن توضيح ذلك من خلال شكل (8.3).

(21) For example, it can be proved that any linear combination of the β's, such as $(\beta_1 - 2\beta_2)$, can be estimated by $(\hat{\beta}_1 - 2\hat{\beta}_2)$, and this estimator is BLUE. For details, see Henri Theil, Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp. 401–402, Note a technical point about the Gauss-Markov theorem: It provides only the sufficient (but not necessary) condition for OLS to be efficient. I am indebted to Michael McAleer of the University of Western Australia for bringing this point to my attention.

في شكل 8.3(a) يتضح أن توزيع المعاينة sampling distribution لتقدير المربعات الصغرى OLS $\hat{\beta}_2$. وهذا التوزيع يوضح احتمالات قيم $\hat{\beta}_2$ في تجارب المعاينة المتكررة repeated sampling experiments (ارجع إلى جدول 1.3) . وللملاءمة سوف نفترض أن $\hat{\beta}_2$ تتوزع توزيعاً متماثلاً symmetrically (وتوجد توضيحات أكثر في الفصل 4) . ومن الشكل يتضح أن توقع $\hat{\beta}_2$ أي $E(\hat{\beta}_2)$ يساوي القيمة الفعلية لـ β_2 . في هذه الحالة، يقال إن التقدير $\hat{\beta}_2$ تقدير غير متحيز unbiased estimator للمعلمة β_2 . في الشكل 8.3(b) يتضح أن توزيع المعاينة للمقدر β_2^* تقدير آخر للمعلمة β_2 التي أمكن الحصول عليها باستخدام طريقة أخرى غير طريقة OLS وللملاءمة سوف نفترض أن β_2^* مثل $\hat{\beta}_2$ أي تقدير غير متحيز unbiased أو بعبارة أخرى $E(\hat{\beta}_2) = E(\beta_2^*) = \beta_2$

وبافتراض أن كلاً من التقديرين β_2^* ، $\hat{\beta}_2$ تقدير خطي linear estimator ، بمعنى أن كلاً منهما دالة خطية في المتغير Y . وهنا يتبقى السؤال أي تقدير من التقديرين β_2^* ، $\hat{\beta}_2$ أفضل ؟

وإجابة هذا السؤال تتضح في شكل 8.3(c) ، حيث يوضح الشكل أنه رغم أن توقع كل من β_2^* ، $\hat{\beta}_2$ يساوي المعلمة β_2 ولكن يتضح من الشكل أيضاً أن قيم التقدير β_2^* أكثر انتشاراً حول β_2 ، في حين أن قيم التقدير $\hat{\beta}_2$ أكثر تكثيفاً حول المعلمة β_2 ، وهذا يعني إحصائياً أن تباين التقدير β_2^* أكبر من تباين التقدير $\hat{\beta}_2$ وبالتالي يكون التقدير الأفضل هو التقدير ذو التباين الأقل، وبالتالي يعتبر التقدير $\hat{\beta}_2$ أفضل . ويسمى التقدير $\hat{\beta}_2$ في هذه الحالة أفضل تقدير خطي غير متحيز best linear unbiased estimator .

ونظرية جاوس ماركوف the Gauss- Markof theorem في فروضها تتضمن التوزيع الاحتمالي للمتغير العشوائي u_i (وبالتالي Y_i وفي الفصل التالي سوف نتناول ذلك) . وطالما أن فروض CLRM تكون متحققة، بالتالي يتم تناول النظرية لجاوس ماركوف . وكنتيجة نحن لانبحث عن تقدير خطي غير متحيز ذي أقل تباين عند استخدام طريقة المربعات الصغرى .

ولكن في حالة عدم توافر الفروض السابق ذكرها ، فإنه لايمكن استخدام نظرية جاوس ماركوف لعدم توافر الفروض التي يثبت عليها . ومثال ذلك بالنسبة لنماذج الانحدار غير الخطية في المعلمات (كما سوف نناقش بالتفصيل في الفصل 14) يمكن الحصول على تقديرات أفضل من تقديرات OLS .

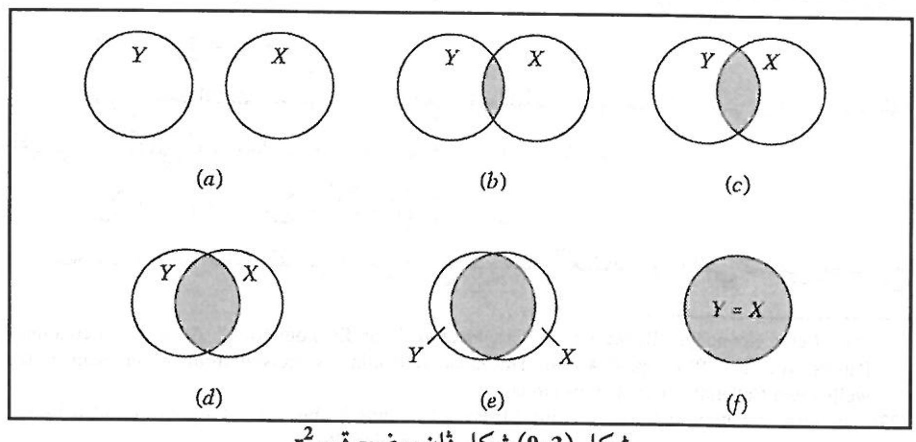
والخصائص الإحصائية statistical properties: التي تم على أساسها المناقشة السابقة، تعتبر خصائص عينة محدودة :finite sample properties . هذه الخصائص تقل عن حجم العينة التي على أساسها تم حساب التقديرات . وفيما بعد سوف نتناول خصائص المقاربة asymptotic properties . هذه الخصائص سوف نتناولها عندما يكون حجم العينة كبيرًا جداً (أي يؤول إلى ما لا نهاية) وملحق A يتناول خصائص التقديرات المحسوبة من عينات بأحجام محددة finite samples وأحجام كبيرة large samples .

5.3 معامل التحديد r^2 .. قياس جودة التوفيق:

THE COEFFICIENT OF DETERMINATION r^2: A MEASURE OF "GOODNESS OF FIT"

وكما سبق، كان اهتمامنا بخصائص تقديرات الانحدار وأخطائها المعيارية Standard errors . وفيما يلي، سوف نتناول جودة توفيق خط الانحدار لمجموعة من البيانات، أو بعبارة أخرى، سوف نتناول هل خط الانحدار يمثل البيانات تمثيلاً جيداً أم لا؟ .

من شكل (1.3) يتضح أن قيم المشاهدات تقع على خط الانحدار، وهنا يعتبر خط الانحدار يمثل البيانات تمثيلاً تاماً perfect ولكن يحدث ذلك في حالات نادرة. وعموماً فإنه يوجد الاختلافات الموجبة $\hat{u}_i$ أو السالبة $\hat{u}_i$ (أو بعبارة أخرى توجد مشاهدات تقع فوق خط الانحدار أو تقع أسفل خط الانحدار) . ويصبح هدفنا هو أن يكون مجموع البواقي حول خط الانحدار أقل مايمكن . ومعامل التحديد r^2 (في حالة متغيرين) أو R^2 (في الانحدار المتعدد) هو مقياس يوضح مدى مواءمة خط انحدار العينة للبيانات .



شكل (3-9) شكل ثان يوضح قيم r^2

وقبل أن نوضح كيف يتم حساب المقياس r^2 . وأشكال ڤان[22] في (9.3) venn diagram توضح القيم المختلفة r^2، كذلك تشير قيم المؤشر إلى النسبة التي يمكن[23] أن يفسر بها المتغير التابع Y عن طريق المتغير المفسر X، أما شكل (b), (c), (d), (e) تشير إلى تزايد نسبة تفسير Y عن طريق X على الترتيب، أما (f) فتشير إلى تطابق X، Y أي عن طريق X يمكن تفسير نسبة 100% للمتغير Y. حيث تتراوح قيم القياس r^2 بين الصفر والواحد أي $0 \leq r^2 \leq 1$.

ولحساب r^2 ، دعنا نعتبر المعادلة:

$$Y_i = \hat{Y}_i + \hat{u}_i \tag{3.6.2}$$

أو معادلة الانحرافات:

$$y_i = \hat{y}_i + \hat{u}_i \tag{1.5.3}$$

وبتربيع طرفي المعادلة (1.5.3) ثم أخذ مجموع الطرفين نحصل على المعادلة التالية:

$$\begin{aligned}\sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i\hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \\ &= \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2\end{aligned} \tag{2.5.3}$$

حيث :

$$\sum \hat{y}_i\hat{u}_i = 0$$

$$\hat{y}_i = \hat{\beta}_2 x_i$$

حيث إن مجاميع المربعات المختلفة في المعادلة (2.5.3) يمكن أن توصف كما يلي:

$$\sum y_i^2 = \sum (Y_i - \bar{Y})^2 =$$

مجموع انحرافات القيم Y عن توقع متوسطات العينات $\bar{Y}$ ، التي يمكن تسميتها بالمجموع ≠ الكلي للمربعات the total sum of squares (TSS) .

$$\sum \hat{y}_i^2 = \sum \left(\hat{Y}_i - \bar{\hat{Y}}\right)^2 = \sum \left(\hat{Y}_i - \bar{Y}\right)^2 = \hat{\beta}_2^2 \sum x_i^2 =$$

مجموع انحرافات تقدير Y عن توقع متوسط العينات $\bar{Y}$ والتي تسمى بمجموع

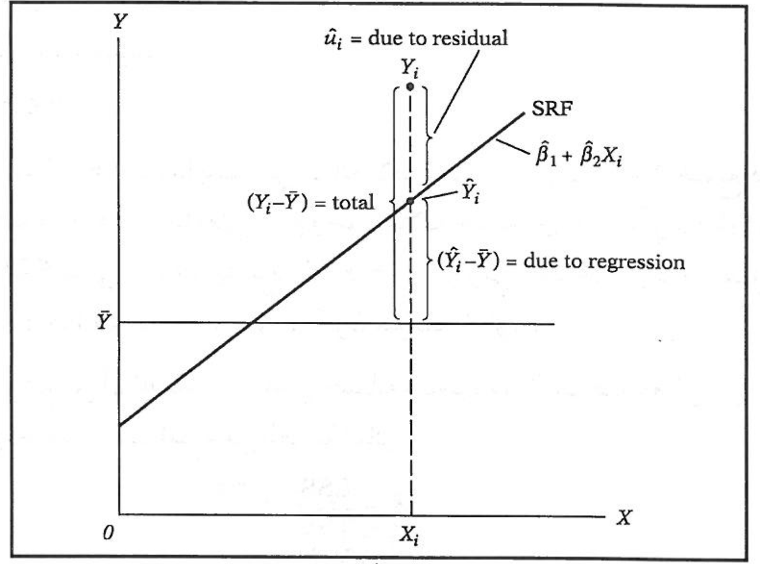
(22) See Peter Kennedy, "Ballentine: A Graphical Aid for Econometrics," Australian Economics Papers, vol. 20, 1981, pp. 414–416. The name Ballentine is dreived from the emblem of the wellknown Ballantine beer with its circles.

(23) The term variation and variance are different. Variance is this sum of squares divided by the apprpriate degrees of freedom. In short, vaiance = variation/df.

مربعات الانحدار (أو بعبارة أخرى الانحرافات التي ترجع إلى المتغيرات المفسرة X)، أو تفسر باستخدام الانحدار، أو مجموع المربعات المفسرة explained sum of squares (ESS). ويسمى المجموع $\sum \hat{u}_i^2$ بالبواقي أو الانحرافات غير المفسرة unexplained sum of squares (RSS)، وهكذا يمكن التعبير عن المعادلة (2.5.3) على النحو التالي:

$$TSS = ESS + RSS \qquad (3.5.3)$$

ومن هنا يتضح أن الانحرافات الكلية total variation في القيم المشاهدة للمتغير Y عن $\bar{Y}$ يمكن تقسيمها إلى جزءين، جزء يرجع إلى الانحدار ($\hat{Y}$: $\bar{Y}$)، والجزء الثاني يرجع إلى البواقي $\hat{u}_i$ كما هو موضح في شكل (10.3).



شكل (3-10) يوضح الأجزاء المختلفة للانحرافات

وبقسمة طرفي المعادلة (3.5.3) على المقدار TSS نحصل على المعادلة التالية:

$$1 = \frac{ESS}{TSS} + \frac{RSS}{TSS}$$
$$= \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} + \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} \qquad (4.5.3)$$

ويعرف r^2 على النحو التالي:

$$r^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{ESS}{TSS} \qquad (5.5.3)$$

أو

$$r^2 = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2} = 1 - \frac{\text{RSS}}{\text{TSS}} \quad \text{(a5.5.3)}$$

ويسمى المقدار r^2 (في حالة العينة) بمعامل التحديد coefficient of determination، ويستخدم كمقياس للحكم على جودة توفيق خط الانحدار. والمقياس r^2 يمثل النسبة المئوية للاختلافات الكلية total variation في المتغير Y التي يمكن تفسيرها عن طريق نموذج الانحدار:

ويتميز r^2 بالخصائص التالية:

1 - مقدار غير سالب.

2 - $0 \leq r^2 \leq 1$

وعندما $r^2 = 1$ فهذا يشير إلى أن العلاقة تامة، بمعنى أن $\hat{Y}_i = Y_i$ لجميع قيم i. كذلك عندما $r^2 = 0$ فهذا يعني عدم وجود علاقة خطية بين Y, X (أي $\hat{\beta}_2 = 0$). وفي هذه الحالة كما في (9.1.3) يتضح أن $\hat{Y}_i = \hat{\beta}_1 = \bar{Y}$. وفي هذه الحالة يكون أفضل تنبؤ best prediction لأي قيمة للمتغير Y تكون متوسط Y أي $\bar{Y}$.

وبالرغم من أن المقياس r^2 يمكن حسابه مباشرة من التعريف المعطى في (3.5.3) أنه يمكن الحصول عليه أسرع من العلاقة التالية:

$$r^2 = \frac{\text{ESS}}{\text{TSS}} = \frac{\sum \hat{y}_i^2}{\sum y_i^2} = \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum y_i^2} = \hat{\beta}_2^2 \left(\frac{\sum x_i^2}{\sum y_i^2}\right) \quad \text{(6.5.3)}$$

وإذا قسمت كل من البسط والمقام في (6.5.3) على حجم العينة n (أو $(n-1)$ إذا كان حجم العينة صغيراً)، نحصل على التالي:

$$r^2 = \hat{\beta}_2^2 \left(\frac{S_x^2}{S_y^2}\right) \quad \text{(7.5.3)}$$

حيث S_y^2 ، S_x^2 يرمزان إلى تباين Y ، تباين X في العينة على الترتيب.

وبما أن $\hat{\beta}_2 = \sum x_i y_i / \sum x_i^2$ في المعادلة (6.5.3) بالتالي يمكن التعبير عن r^2 على النحو التالي:

$$r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2} \qquad (8.5.3)$$

وتعتبر صياغة r^2 على الصورة في العلاقة (8.5.3) أبسط في الحساب، كذلك يمكن وضع كل من ESS، RSS في صورة أبسط للحساب في حالة معرفة r^2 على النحو التالي:

$$\begin{aligned} ESS &= r^2 \cdot TSS \\ &= r^2 \sum y_i^2 \end{aligned} \qquad (9.5.3)$$

$$\begin{aligned} RSS &= TSS - ESS \\ &= TSS(1 - ESS/TSS) \\ &= \sum y_i^2 \cdot (1 - r^2) \end{aligned} \qquad (10.5.3)$$

وبالتالي :

$$\begin{aligned} TSS &= ESS + RSS \\ \sum y_i^2 &= r^2 \sum y_i^2 + (1 - r^2) \sum y_i^2 \end{aligned} \qquad (11.5.3)$$

ومن المقياس r^2 يمكن الحصول على معامل الارتباط r coefficient of correlation الذي سبق الإشارة إليه في الفصل الأول. ومعامل الارتباط r هو مقياس يقيس قوة العلاقة بين المتغيرين Y, X.

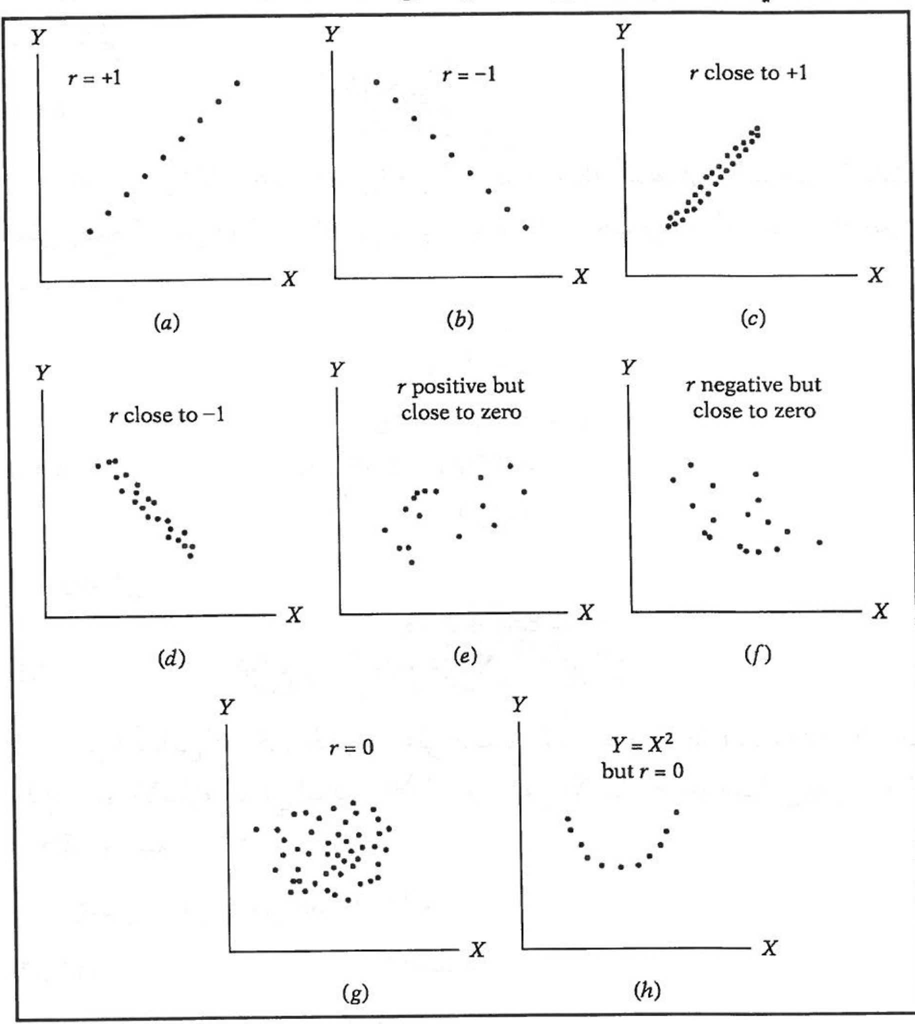
ويمكن حساب r من العلاقة التالية:

$$r = \pm\sqrt{r^2} \qquad (12.5.3)$$

ويمكن صياغة r في الصياغة التالية:

$$\begin{aligned} r &= \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}} \\ &= \frac{n \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{[n \sum X_i^2 - (\sum X_i)^2][n \sum Y_i^2 - (\sum Y_i)^2]}} \end{aligned} \qquad (13.5.3)$$

حيث يعرف Y في (13.5.3) بمعامل ارتباط العينة sample correlation coefficient[24]. وفيما يلي خصائص المقياس r (انظر شكل (3-11)).



شكل (11.3) النماذج المختلفة لقيم معامل الارتباط الخطي

1 - ممكن أن يأخذ قيمة موجبة، عندما تكون العلاقة طردية، ويأخذ قيماً سالبة عندما تكون العلاقة سالبة، ويأخذ القيمة صفر عندما لاتوجد علاقة أو تكون العلاقة غير خطية.

2 - تقع قيمة r بين 1-، 1+ أي: $-1 \leq r \leq +1$

(24) The population correlation coefficient, denoted by ρ , is defined in App. A.

3 – معامل الارتباط ذو طبيعة متماثلة، بمعنى أن معامل الارتباط بين X، $Y(r_{xy})$ يساوي معامل الارتباط بين Y، $X(r_{yx})$.

4 – قيمة r مستقلة عن نقطة الأصل origin والقياس scale، بمعنى أنه إذا كان:

$$X_i^* = aX_i + C\,, \quad Y_i^* = bY_i + d$$

حيث $a > 0$، $b > 0$، c، d فإن،

معامل الارتباط بين X, Y يساوي معامل الارتباط بين X^*, Y^*.

5 – إذا كان X، Y متغيرين مستقلين إحصائياً statistically independent (انظر ملحق A بالنسبة لتعريف الاستقلال الإحصائي)، بمعنى أن معامل الارتباط بين X، Y يساوي صفراً، ولكن إذا كان $r = 0$ فهذا لايعني أنهما مستقلان، أو بعبارة أخرى، مساواة معامل الارتباط بالصفر لايعني بالضرورة استقلال المتغيرات [انظر شكل 11.3(h)].

6 – مقياس خطي يصف العلاقة الخطية، ولايصف العلاقات الأخرى.

7 – بالرغم من أنه مؤشر يصف العلاقة الخطية بين X، Y، ولكنه لايصف العلاقة السببية (أي لايحدد X هي المتغير المفسر، Y المتغير التابع أو العكس) كما سبق توضيحه في الفصل الأول.

ويعتبر r^2 مقياساً أكثر معنوية من r[25].

ويمكن حساب r^2 كمربع لـ r على النحو التالي:

$$r^2 = \frac{\left[\sum(Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})\right]^2}{\sum(Y_i - \bar{Y})^2 \sum(\hat{Y}_i - \bar{Y})^2}$$

أو:

$$r^2 = \frac{\left(\sum y_i \hat{y}_i\right)^2}{\left(\sum y_i^2\right)\left(\sum \hat{y}_i^2\right)} \tag{14.5.3}$$

حيث Y_i تشير إلى القيمة رقم (i) للمتغير Y، $\hat{Y}_i$ تشير إلى تقدير Y_i، كذلك $\bar{\hat{Y}} = \bar{Y}$ تساوي متوسط Y، للإثبات انظر تمرين 15.3. العلاقة (14.5.3) وصف لـ r^2 كمقياس لجودة التوفيق.

(25) In regression modeling the underlying theory will indicate the direction of causality between Y and X, which, in the context of single-equation models, is generally from X to Y.

6.3 مثال رقمي: A NUMERICAL EXAMPLE

وسوف نوضح النظرية الاقتصادية، وذلك بمناقشة دالة الاستهلاك الكنزية Keynesian consumption function السابق تقديمها في المقدمة. بالرجوع إلى تحديد كينز أن قانون علم النفس الرئيسي أن الرجال (السيدات) يعتبران متوسط الاستهلاك يزيد بزيادة الدخل.

جدول (2.3)

Y, $	X, $
70	80
65	100
90	120
95	140
110	160
115	180
120	200
140	220
155	240
150	260

ولكن مقدار الزيادة على الاستهلاك لا يزيد عن الزيادة في الدخل. وهذا يعني أن الاستهلاك الحدي (MPC) أكبر من الصفر ولكن أقل من الواحد. ورغم أن كينز لم يحدد الصياغة الصحيحة للدالة التي تمثل العلاقة بين الاستهلاك والدخل، ولكن للتبسيط افترض أن العلاقة علاقة خطية كما في (2.4.2).

ولكن عند اختبار دالة الاستهلاك الكينزية، نحن سوف نستخدم بيانات العينة في جدول (4.2)، والتي تم إعادة عرضها في جدول (2.3). والبيانات الخام raw data المطلوبة للحصول على تقديرات لمعاملات الانحدار، وأخطائها المعيارية، . . . الخ موضحة بجدول (3.3). ومن البيانات الخام أمكن حساب:

$$\hat{\beta}_1 = 24.4545 \qquad \text{var}(\hat{\beta}_1) = 41.1370 \qquad \text{and} \qquad \text{se}(\hat{\beta}_1) = 6.4138$$
$$\hat{\beta}_2 = 0.5091 \qquad \text{var}(\hat{\beta}_2) = 0.0013 \qquad \text{and} \qquad \text{se}(\hat{\beta}_2) = 0.0357$$
$$\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -0.2172 \qquad \hat{\sigma}^2 = 42.1591 \qquad (1.6.3)$$
$$r^2 = 0.9621 \qquad r = 0.9809 \qquad \text{df} = 8$$

وبالتالي ، يصبح نموذج الانحدار على النحو التالي :

$$\hat{Y}_i = 24.4545 + 0.5091X_i \qquad (2.6.3)$$

وشكل (12.3) يوضح هندسياً النموذج (2.6.3)

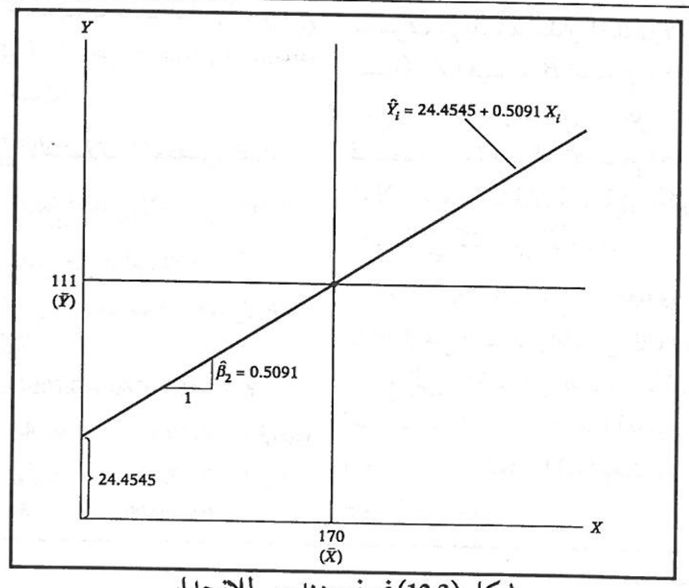
ويمكن تفسير خط الانحدار على النحو التالي : كل نقطة تقع على خط الانحدار تعطي تقـديراً لتوقع Y عند قيـمة معينة لـ X، أي أن $\hat{Y}_i$ تقديـر لـ $E(Y|X_i)$. حيـث قيمة $\hat{\beta}_2 = 0.5091$ ، التي تقيس ميل خط الانحدار، توضح أنه داخل الفترة $\$80 \leq X \leq \$260$ أسبوعياً، عندما تزيد X بدولار واحد، فإن متوسط الإنفاق الاستهلاكي الأسبوعي سوف يزيد بـ 51 سنتًا تقريباً. كذلك قيمة $\hat{\beta}_1 = 24.4545$ تعني أن متوسط الإنفاق الاستهلاكي الأسبوعي عندما يساوي الدخل صفراً يساوي 24.4545 دولار.

جدول (3.3) تقديرات معاملات الانحدار وأخطاؤها

	Y_i (1)	X_i (2)	Y_iX_i (3)	X_i^2 (4)	$x_i = X_i - \bar{X}$ (5)	$y_i = Y_i - \bar{Y}$ (6)	x_i^2 (7)	x_iy_i (8)	$\hat{Y}_i$ (9)	$\hat{u}_i = Y_i - \hat{Y}_i$ (10)	$\hat{Y}_i\hat{u}_i$ (11)
	70	80	5600	6400	−90	−41	8100	3690	65.1818	4.8181	314.0524
	65	100	6500	10000	−70	−46	4900	3220	75.3636	−10.3636	−781.0382
	90	120	10800	14400	−50	−21	2500	1050	85.5454	4.4545	381.0620
	95	140	13300	19600	−30	−16	900	480	95.7272	−0.7272	−69.6128
	110	160	17600	25600	−10	−1	100	10	105.9090	4.0909	433.2631
	115	180	20700	32400	10	4	100	40	116.0909	−1.0909	−126.6434
	120	200	24000	40000	30	9	900	270	125.2727	−6.2727	−792.0708
	140	220	30800	48400	50	29	2500	1450	136.4545	3.5454	483.7858
	155	240	37200	57600	70	44	4900	3080	145.6363	8.3636	1226.4073
	150	260	39000	67600	90	39	8100	3510	156.8181	−6.8181	−1069.2014
Sum	1110	1700	205500	322000	0	0	33000	16800	1109.9995 ≈ 1110.0	0	0.0040 ≈ 0.0
Mean	111	170	nc	nc	0	0	nc	nc	110	0	0

$$\hat{\beta}_2 = \frac{\sum x_iy_i}{\sum x_i^2} = 16{,}800/33{,}000 = 0.5091$$

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X} = 111 - 0.5091(170) = 24.4545$$



شكل (12.3) نموذج هندسي للانحدار

وفي تحليل الانحدار ليست $\hat{\beta}_1$ مقيمة معنوية. ولكن بشكل عام، يجب تفسير قيمة $\hat{\beta}_1$ وفقاً لطبيعة الحالة محل الدراسة. ولكن بشكل عام يمكن تفسير $\hat{\beta}_1$ على أنها متوسط Y عندما لايوجد تأثير للمتغير (أو المتغيرات) X. كذلك قيمة r^2 تساوي 0.9621 تعني أن 96% تقريباً من التغيرات التي تحدث في الإنفاق الاستهلاكي الأسبوعي ترجع إلى التغير في الدخل[26]. ومعامل الارتباط r يساوي 0.9809 تعني أن متغيرات الإنفاق الاستهلاكي والدخل ذو علاقة موجبة قوية. أما الأخطاء المعيارية لتقديرات الانحدار the estimated standard errors سوف تناقش في الفصل (5).

7.3 أمثلة توضيحية : ILLUSTRATIVE EXAMPLES

مثال 3-1: العلاقة بين الدخل والاستهلاك في الولايات المتحدة في الفترة 1982–1996 .

اعتبرنا بيانات الدخل والاستهلاك في جدول (1) في المقدمة. وكما سبق توضيح البيانات في شكل (3-1) باستخدام خط الانحدار في العلاقة (1.3.3) . وهنا سوف توجد تقديرات المربعات الصغرى OLS (النتائج تم الحصول عليها باستخدام البرامج الإحصائية الجاهزة statistical package eviews 3) حيث:

Y : تشير إلى الاستهلاك الشخصي (PCE).

X : تشير إلى الناتج المحلي الإجمالي (GDP).

وتم القياس في عام 1992، والوحدة بالبليون دولار. حيث تم الحصول على المعادلة التالية:

$$\hat{Y}_i = -184.0780 + 0.7064X_i \quad (1.7.3)$$

$$\text{se}(\hat{\beta}_1) = 46.2619 \qquad \text{var}(\hat{\beta}_1) = 2140.1707$$

$$\text{se}(\hat{\beta}_2) = 0.007827 \qquad \text{var}(\hat{\beta}_2) = 0.000061$$

$$\hat{\sigma}^2 = 411.4913 \qquad r^2 = 0.998406$$

المعادلة (1.7.3) تعتبر دالة الاستهلاك الكينزية التجمعية the aggregate Keynesian consumption function (بمعنى الاقتصاد الكلي).

وهذه المعادلة توضح أن الاستهلاك الحدي يساوي 0.71 تقريباً، وهذا يعني أنه عند زيادة الدخل بدولار واحد فإن متوسط الإنفاق الاستهلاكي (PCE) سوف يزيد بمقدار 0.71 دولار (أي 71 سنتًا). أما قيمة $\hat{\beta}_1$ تساوي 184- فهذا يعني أنه عندما يساوي الدخل صفراً فإن قيمة PCE سوف تصبح 184 مليون دولار. وعملياً فإن قيمة $\hat{\beta}_1$ تكون غير معنوية في كثير من الأحيان.

أما قيمة r^2 تساوي 0.9984 فهذا يعني أن 99% من التغيرات في PCE تقريباً ترجع إلى التغير في GDP. وبما أن قيمته تقرب من الواحد فهذا يعني أن خط الانحدار في شكل (1.3) يعتبر توفيقًا جيداً للبيانات.

(26) A formal test of the significance of r^2 will be presented in Chap. 8.

مثال 2.3: إنفاق الطعام في الهند

بالرجوع إلى البيانات في جدول (8.2) بتمرين 15.2 البيانات متعلقة بعينة حجمها 55 أسرة من أسر الريف بالهند. حيث يمثل المتغير التابع إنفاقًا على الطعام ويمثل المتغير المستقل الإنفاق الكلي للأسرة. ووفقاً للبيانات المعطاة نجد أن:

$$\widehat{\text{FoodExp}}_i = 94.2087 + 0.4368\ \text{TotalExp}_i \qquad (3.7.2)$$

$$\text{var}(\hat{\beta}_1) = 2560.9401 \qquad \text{se}(\hat{\beta}_1) = 50.8563$$

$$\text{var}(\hat{\beta}_2) = 0.0061 \qquad \text{se}(\hat{\beta}_2) = 0.0783$$

$$r^2 = 0.3698 \qquad \hat{\sigma}^2 = 4469.6913$$

من العلاقة (2.7.3) نجد أن زيادة الإنفاق الكلي بروبية واحدة سوف يؤدي إلى زيادة الإنفاق على الطعام بمقدار 0.44 روبية تقريباً.

كذلك نجد أن $\hat{\beta}_1 = 94.208$ وهذا يعني أنه عندما يكون الإنفاق الكلي يساوي صفراً فإن متوسط الإنفاق على الطعام يساوي 94.208 روبية. كذلك نجد أن r^2 يساوي 0.37 تقريباً وهذا يعني أن 37% من التغير في الإنفاق على الطعام ترجع إلى التغير في الإنفاق الكلي.

مثال 3.3: العلاقة بين الإيراد والحالة التعليمية

في جدول (6.2)، توجد بيانات عن الإيراد بالساعة، والمستوى التعليمي للفرد بالسنوات، فإذا اعتبرنا أن الإيراد في الساعة Y، والمستوى التعليمي X، فإن:

$$\hat{Y}_i = -0.0144 + 0.7241\ X_i \qquad (3.7.3)$$

$$\text{var}(\hat{\beta}_1) = 0.7649 \quad se(\hat{\beta}_1) = 0.8746$$

$$\text{var}(\hat{\beta}_2) = 0.00483 \quad se(\hat{\beta}_2) = 0.0695$$

$$r^2 = 0.9077 \qquad \hat{\sigma}^2 = 0.8816$$

ومن العلاقة (3.7.3) نجد أن زيادة المستوى التعليمي بسنة دراسية واحدة سوف يؤدي إلى زيادة الإيراد في الساعة بـ 0.72 دولار أي 72 سنتاً تقريباً.

8.3 بعض الملاحظات على تجارب مونت كارلو:
A NOTE ON MONTE CARLO EXPERIMENTS

في هذا الفصل، سوف نوضح أنه تحت فروض CLRM تقديرات المربعات الصغرى تتميز بخصائص معينة تلخص في BLUE. في ملحق هذا الفصل، سوف نثبت هذه الخصائص، ولكن عملياً كيف تحدد أن الخصائص BLUE متحققة في التقديرات؟ فعلى سبيل المثال، كيف نتحقق أن تقديرات المربعات الصغرى غير متحيزة unbiased؟ للإجابة يمكن استخدام مايسمى بتجارب المحاكاة Monte Carlo experiments.

ولتقديم فكرة التحيز، سوف نعتبر PRF على النحو التالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (1.8.3)$$

وفيما يلي خطوات تجربة الـ Monte Carlo:

1 - بافتراض أن $\beta_1 = 20$، $\beta_2 = 0.6$.

2 - اختيار عينة مكونة من 25 فرداً أي $n = 25$.

3 - تحديد قيم معينة لـ X (أي 25 قيمة لـ X).

4 - اختيار 25 قيمة من جدول الأعداد العشوائية random number table أي u_i (وحالياً معظم الحزم الإحصائية يتم استخدامها في توليد الأعداد العشوائية)[28].

5 - بالتعويض في الطرف الأيمن للعلاقة (1.8.3) بقيم X_i، u_i، β_1، β_2 نحصل على قيم Y_i المناظرة. أي نحصل على 25 قيمة لـ Y.

6 - وبأخذ الـ 25 قيمة لـ Y_i التي يتم حسابها في (0)، وكذلك قيم الـ 25 لـ X التي تم تحديدها في (4) يتم تقدير $\hat{\beta}_2$ و $\hat{\beta}_1$ باستخدام طريقة المربعات الصغرى.

7 - بافتراض تكرار التجربة 99 مرة، وفي كل مرة يفترض 25 قيمة لـ X تم حساب 25 قيمة لـ Y وتقدير $\hat{\beta}_1$ و $\hat{\beta}_2$ في كل مرة.

(عملياً هذه التجربة يمكن أن تكرر من 1000 إلى 2000 مرة).

8 - نحسب المتوسط لكل من $\hat{\beta}_2$ و $\hat{\beta}_1$ أي يتم حساب $\bar{\hat{\beta}}_1$ و $\bar{\hat{\beta}}_2$.

9 - فإذا كان:

$$E(\hat{\beta}_2) = \bar{\hat{\beta}}_2 = \beta_2, \quad E(\hat{\beta}_1) = \bar{\hat{\beta}}_1 = \beta_1$$

في هذه الحالة، يقال إن تقديرات المربعات الصغرى تقديرات غير متحيزة.

وهذا النوع من التجارب يستخدم لدراسة الخصائص الإحصائية للتقديرات باستخدام الطرق المختلفة لتقدير معلمات المجتمع.

(28) In practice it is assumed that u_i follows a certain probability distribution, say, normal, with certain parameters (e.g., the mean and variance). Once the values of the parameters are specified, one can easily generate the u_i using statistical packages.

وهذه التجارب تدرس سلوك التقديرات في حالة التجارب ذات العينات صغيرة الحجم ، أو ذات الحجم المحدد finite. وسوف نقدم عديداً من الأمثلة لتجارب الـ Monte Carlo باستخدام التمرينات (انظر تمرين 3-27).

9.3 الملخص والاستنتاجات: SUMMARY AND CONCLUSION

فيما يلي ملخص لأهم النقاط والمفاهيم المقدمة في هذا الفصل:

1 - الإطار الأساسي لتحليل الانحدار هو CLRM.

2 - النموذج CLRM مبني أساساً على مجموعة من الفروض assumptions.

3 - وفقاً للفروض في (2) وباستخدام طريقة المربعات الصغرى، يتم الحصول على تقديرات المربعات الصغرى التي تتميز بمجموعة من الخصائص certain properties التي تتلخص في نظرية جاوس ماركوف the Gauss- Markov theorem والتي تثبت أن هذه التقديرات خطية وغير متحيزة linear unbiased estimators كذلك تقديرات المربعات الصغرى لها أقل تباين minimum variances وباختصار يرمز لها بـ BLUE.

4 - دقة تقديرات OLS تقاس من خلال الأخطاء المعيارية their standard errors. في الفصلين الرابع والخامس، سوف نوضح كيف أن الأخطاء العشوائية أحد الأدوات للحصول على استدلالات حول معلمات المجتمع β population parameters.

5 - جودة توفيق نموذج الانحدار تقاس بمعامل التحديد coefficient of determination r^2. ويوضح هذا المعامل نسبة الاختلافات في المتغير التابع Y التي ترجع إلى الاختلافات في المتغير المفسر X. وأن قيمة r^2 محصورة بين الصفر والواحد.

6 - يرتبط بمعامل التحديد، مقياس آخر مهم هو معامل الارتباط coefficient of correlation r. حيث يقيس معامل الارتباط شدة العلاقة بين المتغير Y، والمتغير X، حيث تنحصر قيمة Y بين -1، $+1$.

7 - والنموذج CLRM هو بناء نظري مجرد مبني على مجموعة من الفروض التي يمكن أن تكون غير واقعية unrealistic. ولكن هذا النموذج المجرد يعتبر مرحلة أولية ضرورية للدراسة والمعرفة. وعند وضع CLRM يمكن اكتشاف مايحدث لو فرض أو أكثر لم يتم تحقيقه.

والجزء الأول من الكتاب، تركز الدراسة على النموذج CLRM. والأجزاء الباقية من الكتاب تتناول كيفية تنقية وتهذيب النموذج CLRM بحيث يتواءم مع الواقع. وجدول (4.3) يوضح رقم الفصل أو الجزء في الكتاب.

جدول (4.3)

الذي يتناول عدم تحقق فرض أو أكثر من فروض CLRM

Assumption number	Type of violation	Where to study?
1	Nonlinearity in parameters	Chapter 14
2	Stochastic regressor(s)	Introduction to Part II
3	Nonzero mean of u_i	Introduction to Part II
4	Heteroscedasticity	Chapter 11
5	Autocorrelated disturbances	Chapter 12
6	Nonzero covariance between disturbances and regressor	Introduction to Part II and Part IV
7	Sample observations less than the number of regressors	Chapter 10
8	Insufficient variability in regressors	Chapter 10
9	Specification bias	Chapters 13, 14
10	Multicollinearity	Chapter 10
11*	Nonnormality of disturbances	Introduction to Part II

**Note:* The assumption that the disturbances u_i are normally distributed is not a part of the CLRM. But more on this in Chapter 4.

EXERCISES تمارين:

أسئلة: Questions

1.3 في العمود الأول (1) من الجدول التالي، توجد ثلاثة فروض تناظرها ثلاث علاقات في العمود الثاني (2) بالجدول. وضح الفروض الأخرى التي على أساسها تثبت العلاقات في العمود (2).

2.3 وفقاً لمرجع Malinvaud (انظر الهامش رقم 10 أسفل صفحة 25)، افترض أن $E(u_i \mid X_i) = 0$ يكون مهماً جداً. لمعرفة ذلك اعتبر الدالة PRR: $Y = \beta_1 + \beta_2 X_i + u_i$. فإذا اعتبرنا الحالتين التاليتين:

(i) $\beta_1 = 0,\ \beta_2 = 0,\ \text{and}\ E(u_i) = 0,$

(ii) $\beta_1 = 1,\ \beta_2 = 0,\ \text{and}\ E(u_i) = (X_i - 1)$

فإذا تم إيجاد التوقع للدالة PRF بالنسبة لـ X. هل تتفق مع Malinvaud بالنسبة للحالتين (i)، (ii)، أن الفرض $E(u_i \mid X_i) = 0$ فرض معنوية.

3.3 اعتبر دالة الانحدار:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$$

اعتبر القيدين:

$$\text{(i) } \sum \hat{u}_i = 0 \quad \text{and} \quad \text{(ii) } \sum \hat{u}_i X_i = 0$$

أوجد التقديرات $\hat{\beta}_1$ و $\hat{\beta}_2$ ثم وضح تطابق هذه التقديرات مع تطابق المربعات الصغرى في المعادلات (6.1.3)، (7.1.3)، هذه الطريقة للحصول على التقديرات تسمى التماثل الرئيسي the analogy principle. اعط مبرراً بديهياً لوضع القيود (i)، (ii). (استرجع فروض النموذج CLRM حول u_i). لاحظ أن التماثل الرياضي لتقديرات المعلمات غير المعلومة يسمى بطريقة العزوم method of moments فيها عزوم العينة (على سبيل المثال متوسط العينة sample mean) تستخدم لتقدير عزوم المجتمع population moments التي تلخص إحصاء statistic التوزيع الاحتمالي probability dist، مثل القيمة المتوقعة expected value والتباين variance.

4.3 وضح أن r^2 المعرفة في (5.5.3) تقع بين الصفر والواحد. يمكن أن تستخدم تباينة Cauchy Schwarz التي تثبت أنه لأي متغيرين X، Y فإن:

$$[E(XY)]^2 \leq E(X^2)E(Y^2)$$

5.3 اعتبر كل من التقديرين $\hat{\beta}_{YX}$، $\hat{\beta}_{XY}$ اللذين يمثلان انحدار Y على X، Y على التوالي، اثبت أن:

$$\hat{\beta}_{YX}\hat{\beta}_{XY} = r^2$$

حيث r تشير إلى معامل الارتباط coefficient of correlation بين X، Y.

6.3 افترض في تمرين 3–5 أن $\hat{\beta}_{YX}\hat{\beta}_{XY} = 1$ هل ذلك يرجع إلى انحدار Y على X أو X على Y؟ وضح ذلك.

7.3 معامل ارتباط سبيرمان Spearman's rank correlation r يعرف على النحو التالي:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

حيث d تساوي الفرق في رتب X، Y المناظرة لنفس المفردات، n تساوي عدد المفردات أو الترتيب للظاهرة.

8.3 اعتبر الصياغات التالية للدالة PRF في متغيرين:

$$\text{Model I:} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

$$\text{Model II:} \quad Y_i = \alpha_1 + \alpha_2 (X_i - \bar{X}) + u_i$$

(أ) أوجد تقديرات المعلمات β_1، σ_1 هل التقديران متطابقان؟ كذلك هل تبايناتهما متطابقة؟

(ب) أوجد تقديرات المعادلتين β_2، σ_2 هل هما متطابقان؟ هل تبايناتهما متطابقة؟

9.3 افترض دالة الانحدار التالية:

$$y_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{u}_i$$

حيث y_i، x_i انحراف كل من Y، X عن $\bar{Y}$ ، $\bar{X}$ على التوالي. ماهي قيمة $\hat{\beta}_1$ ؟ ولماذا؟ وهل $\hat{\beta}_2$ نفس القيمة التي يتم الحصول عليها من المعادلة (6.1.3) ولماذا؟

10.3 اعتبر r_1 تساوي معامل الارتباط بين n زوج من قيم (Y_i, X_i)، كذلك r_2 تساوي معامل الارتباط بين n زوج من القيم $(aX_i + b, cY_i + d)$. حيث a, b, c, d مقادير ثابتة. اثبت أن r_1, r_2، وبالتالي فهذا يؤكد أن معامل الارتباط لايتغير بالنسبة لتغير القياس scale أو تغير نقطة الأصل) origin. (طبق تعريف (1.5.3)).

ملحوظة: العمليات $aX_i, X_i + b$ كذلك $aX_i + b$ تعرف على التوالي كمتغير في القياس، والتغير في نقطة الأصل.

11.3 إذا كان معامل الارتباط r بين المتغيرين X, Y موجبًا، حدد أي العبارات التالية يكون صحيحاً:

(أ) r بين $(-X, -Y)$ يكون أيضاً موجبًا.

(ب) r بين $(-X, Y)$، كذلك (X, Y) كل منهما يكون موجبًا أو سالبًا.

(جـ) كل من معامل الانحدار β_{xy}, β_{yx} يكونان موجبين.

12.3 إذا كانت المتغيرات X_1, X_2, X_3 متغيرات غير مرتبطة uncorrelated variables ولكل منها له نفس الانحراف المعياري standard deviation. وضح أن معامل الارتباط بين $(X_1 + X_2)$، $(X_2 + X_3)$ يساوي $\frac{1}{2}$. لماذا قيمة معامل الارتباط تختلف عن الصفر.

13.3 في علاقة الانحدار التالية:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

إذا فرضنا ضرب كل قيمة من قيم X في مقدار ثابت وليكن (2). هل سوف يؤدي ذلك إلى تغير في البواقي residuals أو توفيق قيم Y؟

14.3 وضح أن العلاقة (14.5.3) تقيس في الواقع معامل التحديد coefficient of determination (طبق تعريف معامل الارتباط r في العلاقة (13.5.3)، وارجع.

$$\sum y_i \hat{y}_i = \sum (\hat{y}_i + \hat{u}_i)\hat{y}_i = \sum \hat{y}_i^2$$

مع الأخذ في الاعتبار العلاقة (6.5.3))

15.3 اشرح مع توضيح السبب أي العبارات التالية تكون صحيحة، وأيها تكون خطأ، وأيها تكون غير محددة uncertain:

(أ) إن معامل الارتباط بين Y، X ينحصر بين 1+ و 1− فهذا يعني أن التغير بين Y، $\text{cov}(Y, X)$ ينحصر أيضاً بين 1+ و 1−.

(ب) إذا كان معامل الارتباط بين المتغيرين يساوي صفراً، فهذا يدل على أنه لاتوجد علاقة relationship بين المتغيرين.

(جـ) إذا كان انحدار Y_i على $\hat{Y}_i$ (بمعنى انحدار القيم الفعلية Y على القيم التقديرية $\hat{Y}$، تقدير المعلمات $\hat{\beta}_1$ و$\hat{\beta}_2$ تساوي صفراً، وواحداً على التوالي.

16.3 إذا فرضنا نموذج الانحدار ليس دالة في المتغير المفسر X ويأخذ الشكل التالي:

$$Y_i = \beta_i + u_i$$

استخدم طريقة OLS لتقدير β_i.

أوجد تباين تقدير β_1 كذلك RSS.

هل تقدير $\hat{\beta}_1$ له دلالة معينة؟

إذا اعتبرنا النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

هل كان من الضروري إضافة X_i إلى النموذج؟

Problems مسائل :

17.3 جدول (5.3) يعطي ترتيب درجات 10 طلاب في امتحاني منتصف الفصل وآخر الفصل الدراسي. احسب معامل ارتباط الرتب لسبيرمان ثم عقب على النتائج.

18.3 توجد علاقة بين معدل التبادل الاسمي nominal exchange rate والأسعار النسبية relative prices. فإذا رمزنا إلى معدل التبادل الاسمي للمارك الألماني GM بالنسبة للدولار الأمريكي $ (GM/$) بالرمز Y. كذلك إذا كان X تشير إلى نسبة الرقم القياسي للاستهلاك في الولايات المتحدة بالنسبة للرقم القياسي للسعر؟ الاستهلاك الألماني. وبالتالي فإن X تشير إلى السعر النسبي للدولتين. وتم جمع بيانات عن Y، X خلال الفترة 1980–1994 فكانت على النحو التالي:

$$\hat{Y}_t = 6.682 - 4.318X_t \qquad r^2 = 0.528$$
$$\text{se} = (1.22)(1.333)$$

(أ) عقب على نموذج الانحدار، ثم فسر قيمة r^2.

(ب) هل يوجد معنى عندما X_i تأخذ قيمًا سالبة؟ وما هو مغزى ذلك في النظرية الاقتصادية؟

(جـ) إذا فرض إعادة تعريف المتغير X على أنه النسبة بين CPI الألماني إلى CPI الأمريكي. هل تتغير إشارة المتغير X، ولماذا؟

19.3 جدول (3–6) يوضح الأرقام القياسية لمخرجات الساعة الواحدة output per hour (X)، وأجر الساعة الواحدة (Y) بالنسبة لقطاع الأعمال وقطاع الأعمال غير الزراعي business and nonfarm business sector في الولايات المتحدة خلال الفترة 1959–1997 . حيث تعتبر سنة الأساس سنة 1982 بمعنى 1980= 100، والأرقام القياسية موسمية (كل 3 شهور).

جدول (5.3) ترتيب درجات الطلاب في الامتحان

	Student									
Rank	A	B	C	D	E	F	G	H	I	J
Midterm	1	3	7	10	9	5	4	8	2	6
Final	3	2	8	7	9	6	5	10	1	4

جدول (6.3) الأرقام القياسية لمخرجات الساعة الواحدة

Year or quarter	Output per hour of all persons[1]		Compensation per hour[2]	
	Business sector	Nonfarm business sector	Business sector	Nonfarm business sector
1959	50.5	54.2	13.1	13.7
1960	51.4	54.8	13.7	14.3
1961	53.2	56.6	14.2	14.8
1962	55.7	59.2	14.8	15.4
1963	57.9	61.2	15.4	15.9
1964	60.6	63.8	16.2	16.7
1965	62.7	65.8	16.8	17.2
1966	65.2	68.0	17.9	18.2
1967	66.6	69.2	18.9	19.3
1968	68.9	71.6	20.5	20.8
1969	69.2	71.7	21.9	22.2
1970	70.6	72.7	23.6	23.8
1971	73.6	75.7	25.1	25.4
1972	76.0	78.3	26.7	27.0
1973	78.4	80.7	29.0	29.2
1974	77.1	79.4	31.8	32.1
1975	79.8	81.6	35.1	35.3
1976	82.5	84.5	38.2	38.4
1977	84.0	85.8	41.2	41.5
1978	84.9	87.0	44.9	45.2
1979	84.5	86.3	49.2	49.5
1980	84.2	86.0	54.5	54.8
1981	85.8	87.0	59.6	60.2
1982	85.3	88.3	64.1	64.6
1983	88.0	89.9	66.8	67.3
1984	90.2	91.4	69.7	70.2
1985	91.7	92.3	73.1	73.4
1986	94.1	94.7	76.8	77.2
1987	94.0	94.5	79.8	80.1
1988	94.7	95.3	83.6	83.7
1989	95.5	95.8	85.9	86.0
1990	96.1	96.3	90.8	90.7
1991	96.7	97.0	95.1	95.1
1992	100.0	100.0	100.0	100.0
1993	100.1	100.1	102.5	102.2
1994	100.7	100.6	104.4	104.2
1995	101.0	101.2	106.8	106.7
1996	103.7	103.7	110.7	110.4
1997	105.4	105.1	114.9	114.5

[1]Output refers to real gross domestic product in the sector.
[2]Wages and salaries of employees plus employers' contributions for social insurance and private benefit plans. Also includes an estimate of wages, salaries, and supplemental payments for the self-employed.
Source: Economic Report of the President, 1999, Table B-49, p. 384.

(أ) ارسم قيم Y ، X في شكلين منفصلين؟

(ب) ماهي النظرية الاقتصادية وراء العلاقة بين X ، Y؟ وهل شكل الانتشار يؤكد النظرية؟

(جـ) قدر باستخدام طريقة OLS علاقة Y على X.

20.3 باستخدام عينة عشوائية حجمها $n=10$، تم الحصول على :

$$\sum Y_i = 1110 \qquad \sum X_i = 1700 \qquad \sum X_iY_i = 205,500$$

$$\sum X_i^2 = 322,000 \qquad \sum Y_i^2 = 132,100$$

وبإعادة اختبار البيانات اتضح التالي :

Y	X
90	120
140	220

بدلاً من :

Y	X
80	110
150	210

ماهو تأثير استبدال القيم Y, X على معامل الارتباط r؟ أوجد القيمة الصحيحة لـ r.

21.3 جدول (7.3) يوضح الأسعار الذهبية gold prices الرقم القياسي لأسعار الاستهلاك consumer price index (CPI)، والرقم القياسي لسعر تبادل الأسهم stock exchange (NYSE) index في الولايات المتحدة خلال الفترة 1977–1991 . الرقم القياسي (NYSE) يتضمن معظم الأسهم في (NYSE).

جدول (7.3) الأسعار الذهبية لرقم قياسي لأسعار الاستهلاك

Year	Price of gold at New York, $ per troy ounce	Consumer Price Index (CPI), 1982–84 = 100	New York Stock Exchange (NYSE) Index, Dec. 31, 1965 = 100
1977	147.98	60.6	53.69
1978	193.44	65.2	53.70
1979	307.62	72.6	58.32
1980	612.51	82.4	68.10
1981	459.61	90.9	74.02
1982	376.01	96.5	68.93
1983	423.83	99.6	92.63

تابع - جدول (7.3)

1984	360.29	103.9	92.46
1985	317.30	107.6	108.90
1986	367.87	109.6	136.00
1987	446.50	113.6	161.70
1988	436.93	118.3	149.91
1989	381.28	124.0	180.02
1990	384.08	130.7	183.46
1991	362.04	136.2	206.33

(أ) ارسم شكل الانتشار scattergram للأسعار الذهبية CPI، والأرقام القياسية NYSE.

(ب) يعتبر الاستثمار investment سياجاً للتضخم inflation إذا كان معدل سعر العائد return على الأقل يحفظ مستوى التضخم. لاختبار هذا الفرض، فإن ذلك يتطلب بناء نماذج ملائمة للعلاقات التالية:

$$\text{Gold price}_t = \beta_1 + \beta_2 \, \text{CPI}_t + u_t$$
$$\text{NYSE index}_t = \beta_1 + \beta_2 \, \text{CPI}_t + u_t$$

3.23 جدول (3-8) يوضح الناتج المحلي الإجمالي gross domestic product (GDP) للولايات المتحدة خلال الفترة 1959-1977.

(أ) ارسم سلسلة زمنية توضح بيانات GDP ومن الرسم قدر GDP في سنة 1992 .

(ب) إذا أشرنا إلى GDP بالرمز Y، والزمن بالرمز X، اعتبر نموذج الـ GDP على النحو التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

أوجد العلاقة التقديرية للنموذج السابق في حالة السعر الجاري للدولار، والسعر الثابت للدولار لـ GDP.

(جـ) ماذا تفسر β_2؟

(د) إذا وجد فرق بين تقدير β_2 في حالة السعر الجاري والسعر الثابت للدولار لـ GDP، اشرح هذا الفرق؟

جدول (8.3) الناتج المحلي الإجمالي للولايات المتحدة

Year	NGDP	RGDP	Year	NGDP	RGDP
1959	507.2000	2210.200	1979	2557.500	4630.600
1960	526.6000	2262.900	1980	2784.200	4615.000
1961	544.8000	2314.300	1981	3115.900	4720.700
1962	585.2000	2454.800	1982	3242.100	4620.300
1963	617.4000	2559.400	1983	3514.500	4803.700
1964	663.0000	2708.400	1984	3902.400	5140.100
1965	719.1000	2881.100	1985	4180.700	5323.500
1966	787.7000	3069.200	1986	4422.200	5487.700
1967	833.6000	3147.200	1987	4692.300	5649.500
1968	910.6000	3293.900	1988	5049.600	5865.200
1969	982.2000	3393.600	1989	5438.700	6062.000
1970	1035.600	3397.600	1990	5743.800	6136.300
1971	1125.400	3510.000	1991	5916.700	6079.400
1972	1237.300	3702.300	1992	6244.400	6244.400
1973	1382.600	3916.300	1993	6558.100	6389.600
1974	1496.900	3891.200	1994	6947.000	6610.700
1975	1630.600	3873.900	1995	7269.600	6761.700
1976	1819.000	4082.900	1996	7661.600	6994.800
1977	2026.900	4273.600	1997	8110.900	7269.800
1978	2291.400	4503.000			

Note: NGDP = nominal GDP (current dollars in billions).
RGDP = real GDP (1992 billions of dollars).
Source: Economic Report of the President, 1999, Tables B-1 and B-2, pp. 326–328.

(هـ) من النتائج السابقة، ماذا تستنتج عن طبيعة التضخم inflation في الولايات المتحدة بالنسبة للفترة التي أخذت فيها العينة؟

23.3 استخدم بيانات جدول (1) في المقدمة وأثبت المعادلة (1.7.3).

24.3 اعتبر بيانات في تمرين 16.2 :

(أ) ارسم شكل انتشار يوضح العلاقة بين درجة السيدات بالنسبة لدرجة الرجال.

(ب) إذا أوضح شكل الانتشار وجود علاقة خطية. قدر نموذج انحدار مناسب.

(جـ) في حالة وجود علاقة: حدد العلاقة السببية بين الدرجتين.

25.3 اعتبر تمرين 23.3 وأحلل درجات الرياضيات math scores بدلاً من درجات اللغة verbal scores.

26.3 اعتبر جدول (2.3)، حيث $\beta_1 = 25, \beta_2 = 0.5$. بافتراض أن u_i يتبع التوزيع المعتاد القياسي بتوقع صفر وتباين 9 أي $U \sim N\ (0.9)$. باستخدام طريقة مونت

كارلو Monte Carlo ولد 100 عينة عشوائية واحسب في كل مرة قيم β_1 و β_2 .

- ارسم هذه التقديرات .
- ما استنتاجك من دراسة مونت كارلو؟ .

ملحوظة : معظم الحزم الإحصائية statistical packages تولد المتغيرات العشوائية لمعظم التوزيعات الاحتمالية .

ملحق A3 — Appendix 3A

1.A3 اشتقاق تقديرات المربعات الصغرى: derivation of least squares estimates

بإجراء التفاضل الجزئي لـ (2.1.3) بالنسبة لـ $\hat{\beta}_1$ و $\hat{\beta}_2$ على التوالي نحصل على :

$$\frac{\partial(\sum \hat{u}_i^2)}{\partial \hat{\beta}_1} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = -2\sum \hat{u}_i \quad (1)$$

$$\frac{\partial(\sum \hat{u}_i^2)}{\partial \hat{\beta}_2} = -2\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)X_i = -2\sum \hat{u}_i X_i \quad (2)$$

وبمساواة (1)، (2) بالصفر، ثم حل المعادلتين، نحصل على التقديرات في المعادلتين (6.1.3)، (7.1.3).

2.A3 خصائص الخطية وعدم التحيز لتقديرات المربعات الصغرى:
linearity and unbiasedness properties of least squares estimators

من العلاقة (8.1.3) نحصل على :

$$\hat{\beta}_2 = \frac{\sum x_i Y_i}{\sum x_i^2} = \sum k_i Y_i \quad (3)$$

حيث :

$$k_i = \frac{x_i}{(\sum x_i^2)}$$

والعلاقة (3) توضح أن $\hat{\beta}_2$ دالة خطية في Y_i أو تقدير خطي linear estimator، k_c تمثل الوزن النسبي لـ Y_i . وتلقائياً نلاحظ الخصائص التالية للأوزان k_i weights :

1 - بما أن X_i غير عشوائية nonstochastic، بالتالي فإن k_i غير عشوائية أيضاً .

2 - $\Sigma k_i = 0$

3- $\sum k_i^2 = 1/\sum x_i^2$

4- $\sum k_i x_i = \sum k_i X_i = 1$

ويمكن إثبات هذه الخصائص مباشرة من تعريف k_i، فعلى سبيل المثال:

$$\sum k_i = \sum \left(\frac{x_i}{\sum x_i^2}\right) = \frac{1}{\sum x_i^2} \sum x_i,$$

وبما أن $\sum x_i$ تشـير إلى مجمـوع انحـرافـات القـيم عن الوسط الحسـابـي (حـيث $x_i = X_i - \bar{X}$) بالتالي فإن $\sum x_i = 0$ وبالتالي فإن:

$$\sum k_i = \frac{0}{\sum x_i^2} = 0$$

وبالتعويض في الدالة PRF حيث:

$$\text{PRF} \quad Y_i = \beta_1 + \beta_2 X_i + u_i$$

في (3) نحصل على التالي:

$$\begin{aligned} \hat{\beta}_2 &= \sum k_i(\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_1 \sum k_i + \beta_2 \sum k_i X_i + \sum k_i u_i \\ &= \beta_2 + \sum k_i u_i \end{aligned} \tag{4}$$

وباستخدام خصائص k_i السابقة وبأخذ توقع لطرفي العلاقة (4) نجد أن:

$$\begin{aligned} E(\hat{\beta}_2) &= \beta_2 + \sum k_i E(u_i) \\ &= \beta_2 \end{aligned} \tag{5}$$

حيث $E(u_i) = 0$ من فروض المتعلقة بالمتغير u_i. ومن (5) فإن $\hat{\beta}_2$ تقدير غير متحيز للمعلمة β_2. وبنفس الأسلوب يمكن إثبات أن $\hat{\beta}_1$ تقدير غير متحيز للمعلمة β_1.

3.A3 التباينات والأخطاء المعيارية لتقديرات المربعات الصغرى

Variance and standard errors of least squares estimators

وبما أن تباين التقدير $\hat{\beta}_2$ على النحو:

$$\text{var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2$$

حيث:

$$E(\hat{\beta}_2) = \beta_2 \tag{6}$$

بالتالي:

$$\begin{aligned} \text{var}(\hat{\beta}_2) &= E\left(\sum k_i u_i\right)^2 \\ &= E\left(k_1^2 u_1^2 + k_2^2 u_2^2 + \cdots + k_n^2 u_n^2 + 2k_1 k_2 u_1 u_2 + \cdots + 2k_{n-1} k_n u_{n-1} u_n\right) \end{aligned}$$

وبافتراض أن :

$$E(u_i^2) = \sigma^2 , \quad E(u_i u_j) = 0 , \quad i \neq j$$

وبالتالي فإن :

$$\begin{aligned} \operatorname{var}(\hat{\beta}_2) &= \sigma^2 \sum k_i^2 \\ &= \frac{\sigma^2}{\sum x_i^2} \\ &= \text{Eq. } (3.3.1) \end{aligned} \tag{7}$$

وبنفس الأسلوب يمكن الحصول على تباين $\hat{\beta}_1$.

4.A3 التغاير بين $\hat{\beta}_1$ و $\hat{\beta}_2$: Covariance between $\hat{\beta}_1 , \hat{\beta}_2$

من تعريف التغاير

$$\begin{aligned} \operatorname{cov}(\hat{\beta}_1, \hat{\beta}_2) &= E\{[\hat{\beta}_1 - E(\hat{\beta}_1)][\hat{\beta}_2 - E(\hat{\beta}_2)]\} \\ &= E(\hat{\beta}_1 - \beta_1)(\hat{\beta}_2 - \beta_2) \\ &= -\bar{X} E(\hat{\beta}_2 - \beta_2)^2 \\ &= -\bar{X} \operatorname{var}(\hat{\beta}_2) \\ &= \text{Eq. } (3.3.9) \end{aligned} \tag{8}$$

5.A3 تقدير المربعات الصغرى للتباين : The least squares estimator of σ^2

بالرجوع إلى :

$$Y_i = \beta_1 + \beta_2 X_i + u_i \tag{9}$$

كذلك :

$$\bar{Y} = \beta_1 + \beta_2 \bar{X} + \bar{u} \tag{10}$$

بطرح (9) من (10) نجد أن :

$$y_i = \beta_2 x_i + (u_i - \bar{u}) \tag{11}$$

كذلك :

$$\hat{u}_i = y_i - \hat{\beta}_2 x_i \tag{12}$$

بالتعويض بـ (11) في (12) نجد أن :

$$\hat{u}_i = \beta_2 x_i + (u_i - \bar{u}) - \hat{\beta}_2 x_i \tag{13}$$

بتربيع طرفي العلاقة السابقة ثم أخذ المجموع للطرفين نحصل على :

$$\sum \hat{u}_i^2 = (\hat{\beta}_2 - \beta_2)^2 \sum x_i^2 + \sum (u_i - \bar{u})^2 - 2(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u}) \quad (14)$$

وبإيجاد التوقع للطرفين نحصل على مايلي :

$$\begin{aligned} E\left(\sum \hat{u}_i^2\right) &= \sum x_i^2 E(\hat{\beta}_2 - \beta_2)^2 + E\left[\sum (u_i - \bar{u})^2\right] - 2E\left[(\hat{\beta}_2 - \beta_2) \sum x_i (u_i - \bar{u})\right] \\ &= \sum x_i^2 \operatorname{var}(\hat{\beta}_2) + (n-1) \operatorname{var}(u_i) - 2E\left[\sum k_i u_i (x_i u_i)\right] \\ &= \sigma^2 + (n-1)\sigma^2 - 2E\left[\sum k_i x_i u_i^2\right] \\ &= \sigma^2 + (n-1)\sigma^2 - 2\sigma^2 \\ &= (n-2)\sigma^2 \end{aligned} \quad (15)$$

باستخدام الطرف الأيمن للمعادلة السابقة في (3)، (4) نجد أن :

$$\begin{aligned} E\sum (u_i - \bar{u})^2 &= E\left[\sum u_i^2 - n\bar{u}^2\right] \\ &= E\left[\sum u_i^2 - n\left(\frac{\sum u_i}{n}\right)^2\right] \\ &= E\left[\sum u_i^2 - \frac{1}{n}\sum (u_i^2)\right] \\ &= n\sigma^2 - \frac{n}{n}\sigma^2 = (n-1)\sigma^2 \end{aligned}$$

ومما سبق يتضح أن u_i غير مرتبطة، وأن تباين u_i يساوي σ^2.

من (15) نجد أن

$$E\left(\sum \hat{u}_i^2\right) = (n-2)\sigma^2 \quad (16)$$

لذا نعرف :

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \quad (17)$$

وبالتالي فإن :

$$E(\hat{\sigma}^2) = \frac{1}{n-2} E\left(\sum \hat{u}_i^2\right) = \sigma^2 \quad (18)$$

وبالتالي فإن σ^2 تقدير غير متحيز للمعلمة σ^2.

6.A3 خاصية أقل تباين لتقديرات المربعات الصغرى :

Minimum- variance property of least squares estimators

يتضح من ملحق 2A3 الفصل(3) أن تقديرات المربعات الصغرى $\hat{\beta}_2$ خطية، وأيضاً غير متحيزة (وذلك ينطبق أيضاً على $\hat{\beta}_1$. وهنا سوف نوضح أن هذه

التقديرات أيضاً ذات أقل تباين بالنسبة لفئة التقديرات الخطية غير المتحيزة. إذا اعتبرنا تقـدير المربعـات الصغرى $\hat{\beta}_2$

$$\hat{\beta}_2 = \sum k_i Y_i$$

حيث :

$$k_i = \frac{X_i - \bar{X}}{\sum(X_i - \bar{X})^2} = \frac{x_i}{\sum x_i^2} \tag{19}$$

(انظر ملحق 2.A3)

ويتضح أن $\hat{\beta}_2$ متوسط مرجح weighted average لـ Y_i بالوزن من k_i.

إذا فرضنا تقديراً خطياً غير متحيز آخر للمعلمة β_2 وليكن β_2^* حيث :

$$\beta_2^* = \sum w_i Y_i \tag{20}$$

حيث w_i أوزان لـ Y_i وليست بالضرورة تساوي k_i لجميع قيم i.

$$\begin{aligned} E(\beta_2^*) &= \sum w_i E(Y_i) \\ &= \sum w_i(\beta_1 + \beta_2 X_i) \\ &= \beta_1 \sum w_i + \beta_2 \sum w_i X_i \end{aligned} \tag{21}$$

وبما أن β_2^* تقدير غير متحيز بالتالي فإن :

$$\sum w_i = 0 \tag{22}$$

$$\sum w_i X_i = 1 \tag{23}$$

كذلك يمكن إيجاد تباين β_2^* على النحو التالي :

$$\begin{aligned} \text{var}(\beta_2^*) &= \text{var} \sum w_i Y_i \\ &= \sum w_i^2 \, \text{var}\, Y_i \qquad [\textit{Note:} \, \text{var}\, Y_i = \text{var}\, u_i = \sigma^2] \\ &= \sigma^2 \sum w_i^2 \qquad [\textit{Note:} \, \text{cov}(Y_i, Y_j) = 0 \, (i \neq j)] \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} + \frac{x_i}{\sum x_i^2} \right)^2 \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \frac{\sum x_i^2}{(\sum x_i^2)^2} + 2\sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right) \left(\frac{x_i}{\sum x_i^2} \right) \\ &= \sigma^2 \sum \left(w_i - \frac{x_i}{\sum x_i^2} \right)^2 + \sigma^2 \left(\frac{1}{\sum x_i^2} \right) \end{aligned} \tag{24}$$

وبما أن الحد الأخير في الطرف الأيمن للعلاقة (24) مقدار ثابت، وبالتالي فإن أقل قيمة لتباين β_2^* يصبح المقدار $\sigma^2(1/\sum x_i^2)$.

وهذا يؤدي إلى أن $w_i = k_i$، وبالتالي $\hat{\beta}_2$ تقدير خطي غير متحيز ذو أقل تباين. وبنفس الطريقة يمكن إثبات أن تقدير المربعات الصغرى $\hat{\beta}_1$ ذو أقل تباين في فئة التقديرات الخطية غير المتحيزة للمعلمة β_1 .

7.A3 اتساق تقديرات المربعات الصغرى : consistency of least-squares estimators

في هذا الفصل، سوف نوضح أن نموذج الانحدار الخطي التقليدي، وتقدير المربعات الصغرى الخطية غير المتحيزة (كفء efficient) في أي حجم للعينة، بمعنى سواء كان حجم العينة صغيراً أو كبيراً. ولكن أحياناً كما ناقشنا في ملحق A، أن التقدير قد لايحقق خاصية أو أكثر من الخصائص المرغوبة في حالة العينات ذات الحجم الصغير.

وكلما زاد حجم العينة كلما زاد عدد الخصائص المرغوبة في حالة العينات ذات الحجم الصغير. وكلما زاد حجم العينة كلما زاد عدد الخصائص المرغوب فيها في التقدير، والتي تحقق في تقديرات العينات ذات الحجم الكبير، وتسمى هذه الخصائص بخصائص التقارب asymptotic properties .

في هذا الملحق، سوف نتناول خاصية واحدة من خصائص التقارب، وهي خاصية الاتساق consistency، وكما أوضحنا في ملحق A أن تقدير المربعات الصغرى $\hat{\beta}_2$ خطي وغير متحيز. وفي هذا الفصل، سوف نوضح أن $\hat{\beta}_2$ تقدير متسق أيضاً. وفيما يلي، سوف نوضح نهاية تباين $\hat{\beta}_2$ تؤول إلى الصفر عندما يؤول حجم العينة إلى مالانهاية، وذلك عندما يكون التقدير $\hat{\beta}_2$ تقديراً غير متحيز، وذلك على النحو التالي:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} = \frac{\sigma^2/n}{\sum x_i^2/n} \quad (26)$$

$$\underbrace{\lim \operatorname{var}(\hat{\beta}_2)}_{n\to\infty} = \underbrace{\lim \left(\frac{\sigma^2/n}{\sum x_i^2/n}\right)}_{n\to\infty} = 0 \quad (27)$$

وبالتالي، يصبح التقدير $\hat{\beta}_2$ تقديراً متسقًا للمعلمة β_2 وبنفس الطريقة يمكن إثبات أن $\hat{\beta}_1$ تقدير متسق أيضاً. ومن هنا نخلص إلى أن تقديرات المربعات الصغرى تقديرات متسقة عندما يتزايد حجم العينة.

الفصل الرابع

نموذج الانحدار الخطي الطبيعي التقليدي

CLASSICAL NORMAL LINEAR REGRESSION MODEL (CNLRM)

تنقسم النظرية التقليدية classical theory of statistical inference إلى فرعين أساسيين هما :

1 - التقديرات : estimation

2 - اختبارات الفروض : hypothesis testing

وفي هذا الفصل ، سوف نتناول بالدراسة التقديرات estimation لمعلمات نموذج الانحدار الخطي (نموذج متغيرين) . سبق أن تناولنا طريقة المربعات الصغرى العادية OLS ، وحصلنا على تقديرات للمعلمات $\sigma^2, \beta_2, \beta_1$. وذلك تحت فروض نموذج الانحدار الخطي التقليدي (CLRM) . حيث أمكن الحصول على التقديرات $\hat{\sigma}^2, \hat{\beta}_2, \hat{\beta}_1$ التي لها خصائص إحصائية مهمة مثل عدم التحيز unbiasedness ، أقل تباين minimum variance ، ... الخ (راجع خاصية BLUE) . مع ملاحظة أن قيم هذه التقديرات التي تم الحصول عليها من بيانات عينة تختلف باختلاف مفردات العينة . وبالتالي ، فإن تقديرات العينة تعتبر متغيرات عشوائية random variables .

وتعتبر **موضوعات التقدير نصف موضوعات الاستدلال الإحصائي** ، والنصف الآخر هوموضوعات اختبارات الفروض . وبالرجوع إلى تحليل الانحدار ، نجد أن هدفنا ليس فقط إيجاد تقدير دالة الانحدار من العينة sample regression function (SRF) ، ولكن تقدير دالة انحدار المجتمع population regression function (PRF) ، كما سبق الإشارة لذلك في الفصل الثاني .

وبالتالي ، نحاول دراسة كيفية اقتراب $\hat{\beta}_1$ من المعلمة الحقيقية β_1 ، أو كيفية اقتراب $\hat{\sigma}^2$ من المعلمة الحقيقية σ^2 . وعلى سبيل المثال ، في مثال 2.3 تم تقدير الدالة (SRF) في المعادلة (2.7.3) حيث تم التقدير باستخدام بيانات عينة مكونة من 55 أسرة ، ولكن يبقى السؤال كيف يمكن تقدير (MPC) أو القيمة 0.4368 للحصول على (MPC) للمجتمع المسحوب منه العينة؟ ولتحديد علاقة تقدير (MPC) من العينة بـ (MPC) في المجتمع لذلك ، بما أن كل تقدير من التقديرات $\hat{\sigma}^2, \hat{\beta}_2, \hat{\beta}_1$ يعتبر متغيراً عشوائيًا random variables .

ولتحديد العلاقة بين تقدير (MPC) من العينة بتقدير (MPC) في المجتمع ، لابد أولاً من الإلمام بالتوزيعات الاحتمالية للمتغيرات العشوائية $\hat{\sigma}^2, \hat{\beta}_2, \hat{\beta}_1$ فكما ذكرنا من قبل أن التقديرات $\hat{\sigma}^2, \hat{\beta}_2, \hat{\beta}_1$ متغيرات عشوائية random variables . فبدون معرفة التوزيعات الاحتمالية للمتغيرات $\hat{\sigma}^2, \hat{\beta}_2, \hat{\beta}_1$ لا يمكن معرفة العلاقة بين تقديرات معلمات المجتمع ، وتقديرات معلمات العينة .

1.4 التوزيع الاحتمالي للمتغير العشوائي (u_i) :
THE PROBABILITY DISTRIBUTION OF DISTURBANCES u_i

لإيجاد التوزيع الاحتمالي لتقديرات المربعات الصغرى OLS ، سوف نتبع مايلي :

سوف نتناول التقدير $\hat{\beta}_2$ ، فمن الملحق 2.A3 نجد أن :

$$\hat{\beta}_2 = \sum k_i Y_i \qquad \textbf{(1.1.4)}$$

حيث $k_i = \frac{x_i}{\sum x_i^2}$. وحيث إن قيم X قيم محددة fixed أي غير عشوائية nonstochastic ، وذلك من فروض تحليل الانحدار الشرطي conditional regression analysis ، أي تكون القيم X_i قيمًا محددة . من المعادلة (1.1.4) نجد أن $\hat{\beta}_2$ دالة خطية في Y_i ، حيث Y_i متغير عشوائي من الفروض . حيث إن $Y_i = \beta_1 + \beta_2 X_i + u_i$ ، وبالتالي يمكن كتابة المعادلة (1.1.4) على النحو التالي :

$$\hat{\beta}_2 = \sum k_i(\beta_1 + \beta_2 X_i + u_i) \qquad \textbf{(2.1.4)}$$

وبما أن k_i ، X_i قيم محددة ، بالتالي فإن $\hat{\beta}_2$ عبارة عن دالة خطية في متغير عشوائي u_i .

وبالتالي فالتوزيع الاحتمالي لـ $\hat{\beta}_2$ (وبالمثل $\hat{\beta}_2$) سوف يعتمد على افتراض التوزيع الاحتمالي لـ u_i . وبالتالي معرفة التوزيع الاحتمالي لتقديرات المربعات الصغرى OLS يعتبر الأساس في أي استنتاج حول معلمات المجتمع β_1 و β_2 . حيث تلعب طبيعة التوزيع الاحتمالي للمتغير u_i دوراً بالغ الأهمية في اختبارات الفروض hypothesis testing .

وبما أن طريقة المربعات الصغرى لا تضع أي فروض حول الطبيعة الاحتمالية probabilistic nature للمتغير u_i وهذا يساعد في إيجاد الدالة (PRF) من الدالة (SRF) . كما هو مثبت في نظرية جاوس ماركوف Gauss- Marcov theorem .

وعادة يفترض أن المتغير العشوائي u يتبع التوزيع الطبيعي (المعتاد normal distribution . وبإضافة الفرض الطبيعي normaility assumption للمتغير u إلى باقي فروض نموذج الانحدار الخطي التقليدي classical linear regression model (CLRM) . التي سبق مناقشتها في الفصل الثالث ، نحصل على مايسمى بنموذج الانحدار الخطي الطبيعي التقليدي the classical normal linear regression model (CNLRM) .

2.4 افتراض التوزيع الطبيعي للمتغير u_i :
THE NORMALITY ASSUMPTION FOR u_i

يفترض نموذج الانحدار الخطي الطبيعي التقليدي أن كل متغير من المتغيرات u_i يتبع التوزيع الطبيعي (المعتاد) بتوقع صفر وتباين σ^2 أي أن :

$$E(u_i) = 0 \quad (1.2.4)$$

$$\mathrm{Var}(u_i) = E[u_i - E(u_i)]^2 = E(u_i^2) = \sigma^2 \quad (2.2.4)$$

$$\mathrm{cov}(u_i, u_j): \quad E\{[(u_i - E(u_i)][u_j - E(u_j)]\} = E(u_i\, u_j) = 0 \quad i \neq j \quad (3.2.4)$$

وللاختصار يشار إلى هذه الفروض للمتغير u_i بالآتي $u_i \sim N(0, \sigma^2)$ وتقرأ "المتغير u_i متغير يتبع التوزيع الطبيعي بتوقع صفر وتباين σ^2" .

$$u_i \sim N(0, \sigma^2) \quad (4.2.4)$$

وكما أوضحنا في ملحق A ، أن افتراض التوزيع الطبيعي للمتغير u_i كما في (4.2.4) يعني أن المتغيرات $i = 1, 2, ..., \text{nbu}_i$ متغيرات مستقلة independent variables . وبالتالي تعاد كتابة المعادلة (4.2.4) على النحو :

$$u_i \sim \mathrm{NID}(0, \sigma^2) \qquad (5.2.4)$$

وتقرأ " كل متغير من المتغيرات u_i ، $i = 1, 2, ..., n$ متغيرات عشوائية مستقلة لكل منها التوزيع الطبيعي بتوقع صفر وتباين σ^2 .

لماذا افتراض التوزيع الطبيعي : Why the normality assumption

يستخدم افتراض التوزيع الطبيعي للمتغيرات u_i ، $i = 1, 2, ..., n$ للأسباب التالية :

1 – وكما ذكر سابقاً في الفصل 5.2 أن المتغير u_i هو عبارة عن التأثير التجميعي لعدد كبير من المتغيرات المستقلة على المتغير التابع ، حيث لم يوضح ذلك بالتفصيل . ونظراً لوجود نظرية النزعة المركزية central limit theorem (CLT) في الإحصاء (انظر التفاصيل بملحق A) ، أنه يمكن إثبات أن مجموع المتغيرات العشوائية المستقلة المتماثلة identically التي لها التوزيع المعتاد عبارة عن متغير عشوائي له توزيع يؤول إلى التوزيع الطبيعي tends to a normal distribution علماً بأن هذه المتغيرات تتزايد زيادة غير محدودة indefinitely[1] . وتمدنا بطريقة المركزية (CLT) بالتوزيع النظري لمجموع المتغيرات u_i في حالة إذا كان $u_i \sim N(0, \sigma^2)$.

2 – مع افتراض التوزيع الاحتمالي الطبيعي لكل متغير من المتغيرات u_i ، وباستخدام نظرية النزعة المركزية (CLT) نجد أن مجموع المتغيرات u_i ، $i = 1, 2, ..., n$ يتبع التوزيع الطبيعي ، سواء كانت المتغيرات u_i مستقلة independent أو غير مستقلة dependent[2] .

3 – مع افتراض التوزيع الطبيعي لكل متغير من المتغيرات u_i ، يمكن اشتقاق بسهولة التوزيع الاحتمالي لتقديرات المربعات الصغرى OLS كما هو موضح في

---

(1) For a relatively simple and straightforward discussion of this theorem, see Sheldon M. Ross, Introduction, to Probability and Statistics for Engineers and Scientists, 2d ed., Harcourt Academic Press, New York, 2000, pp. 193-194. One exception to the theorem is the Cauchy distribution, which has no mean or higher moments. See M. G. Kendall and A. Stuart, The Advanced Theory of Statistics, Charles Griffin & Co., London, 1960. vol. 1, pp. 248–249.

(2) For the various forms of the CLT, see Harald Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, N.J., 1946, Chap. 17.

ملحق A . وكما سبق أن ذكرنا ، أن إحدي الخصائص للتوزيع الطبيعي ، أن أي دالة خطية في متغيرات تتبع التوزيع الطبيعي تكون متغيراً طبيعيًا أيضاً (أي يتبع التوزيع الطبيعي أيضاً) .

وكما سبق أن أوضحنا ، أن تقديرات المربعات الصغرى $\hat{\beta}_1$ و$\hat{\beta}_2$ دوال خطية في u_i . وبالتالي فإن $\hat{\beta}_1$ و $\hat{\beta}_2$ تتبع كل منهما التوزيع الطبيعي أيضاً . وتعتبر هذه الخاصية بالغة الأهمية في إجراء اختبارات الفروض hypothesis testing .

4 - التوزيع الطبيعي يعتبر توزيعًا بسيطًا له معلمتان فقط (التوقع mean والتباين variance) ، بالإضافة أن خصائصه النظرية معروفة بالتفصيل في الإحصاء الرياضي mathematical statistics ، بالإضافة أن كثيراً من الظواهر تتبع التوزيع الطبيعي .

5 - أخيراً ، في حالة التعامل مع العينات صغيرة الحجم (ويمكن اعتبار العينات صغيرة الحجم إذا كان عدد المشاهدات أقل من 100 مشاهدة) ، افتراض التوزيع الطبيعي لـ u_i تلعب دوراً مهمًا في اشتقاق التوزيعات الاحتمالية الصحيحة exact probability distributions لتقديرات المربعات الصغرى . بالإضافة إلى أنها تمكننا من استخدام الاختبارات الإحصائية (t, F, X_2 statistical tests) . (يمكن الرجوع إلى ملحق A لمناقشة خصائص التوزيعات t, F, X_2) . وكما سوف نوضح فيما بعد ، أنه عندما يكون حجم العينة كبيراً ، فإنه يمكن التوسع في استخدام افتراض التوزيع الطبيعي .

ملحوظة : ورغم أننا تناولنا افتراض التوزيع الطبيعي وأسباب استخدامه ، وأنه يكون ملائمًا بالنسبة لأحجام العينات الصغيرة .

ولكن سوف نوضح فيما بعد أنه في بعض الحالات لايعتبر افتراض التوزيع الطبيعي افتراضًا ملائمًا .

3.4 خصائص تقديرات المربعات الصغرى وفقًا افتراض التوزيع الطبيعي: PROPERTIES OF OLS ESTIMATORS UNDER THE NORMALITY ASSUMPTION

وعندما يكون المتغير u_i يتبع التوزيع المعتاد كما في العلاقة (5.2.4) فإن الخصائص التالية تتوافر في تقديرات المربعات الصغرى (انظر ملحق : A)

1 - تعتبر تقديرات المربعات الصغرى تقديرات غير متحيزة unbiased .

2 - تقديرات ذات أقل تباين minimum variance وبدمج الخاصية 1 مع 2 تسمى التقديرات في هذه الحالة تقديرات كفء efficient estimators .

3 - أيضاً تعتبر تقديرات المربعات الصغرى تقديرات متسقة consistency حيث نجد أن قيمة التقدير تقترب من قيمة المعلمة الفعلية في المجتمع عندما يؤول حجم العينة إلى ما لا نهاية .

4 - $\hat{\beta}_1$ (يعتبر دالة خطية في u_i) متغير يتبع التوزيع الطبيعي (المعتاد) بحيث :

$$\text{Mean:} \quad E(\hat{\beta}_1) = \beta_1 \tag{1.3.4}$$

$$\text{var}(\hat{\beta}_1): \quad \sigma^2_{\hat{\beta}_1} = \frac{\sum X_i^2}{n\sum x_i^2}\sigma^2 \qquad = (3.3.3)\ (2.3.4)$$

أو بعبارة أخرى :

$$\hat{\beta}_1 \sim N(\beta_1, \sigma^2_{\hat{\beta}_1})$$

وباستخدام خصائص التوزيع الطبيعي ، إذا عرفنا المتغير Z بحيث :

$$Z = \frac{\hat{\beta}_1 - \beta_1}{\sigma_{\hat{\beta}_1}} \tag{3.3.4}$$

حيث يتغير المتغير Z التوزيع المعتاد القياسي standard normal distribution حيث توقع Z يساوي صفرًا ، وتباين يساوي واحدًا أي أن :

$$Z \sim N(0, 1)$$

5 - $\hat{\beta}_2$ (دالة خطية في u_i) يتبع التوزيع المعتاد بحيث :

$$\text{Mean:} \quad E(\hat{\beta}_2) = \beta_2 \tag{4.3.4}$$

$$\text{var}(\hat{\beta}_2): \quad \sigma^2_{\hat{\beta}_2} = \frac{\sigma^2}{\sum x_i^2} \quad = (1.3.3)\ (5.3.4)$$

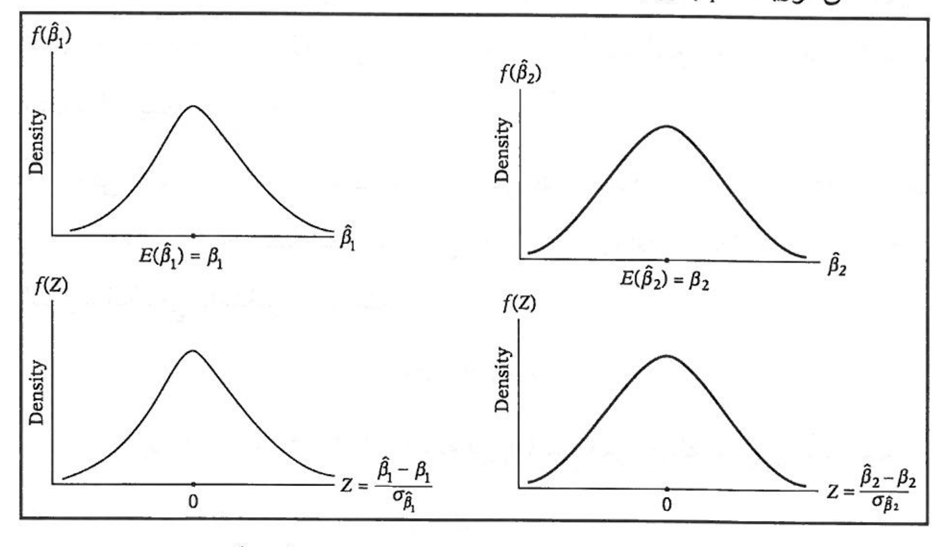
أو بعبارة أخرى :

$$\hat{\beta}_2 \sim N(\beta_2, \sigma^2_{\hat{\beta}_2})$$

وبالتالي فإن :

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\sigma_{\hat{\beta}_2}} \tag{6.3.4}$$

حيث Z يتبع التوزيع المعتاد القياسي أيضاً . وشكل (1.4) التالي يوضح خصائص توزيعات $\hat{\beta}_1$ و $\hat{\beta}_2$.



شكل (1.4) التوزيعان الاحتماليان للتقديرين $\hat{\beta}_1$ و $\hat{\beta}_2$

6 – ويعتبر المتغير $(n-2)(\hat{\sigma}^2/\sigma^2)$ متغيراً يتبع توزيع مربع كاينز X^2 بدرجات حرية[3] $(n-2)$. ومعرفة التوزيع الاحتمالي لهذا المتغير يعتبر ذا أهمية في إجراء استدلال عن المعلمة σ^2 باستخدام التقدير $\hat{\sigma}^2$ (توزيع مربع كاينز X^2 وخصائصه تناقش في ملحق A) .

(3) The proof of this statement is slightly involved. An accessible source for the proof is Robert V. Hogg and Allen T. Craig, Introduction to Mathematical Statistics, 2d ed., Macmillan, New York, 1965, p. 144.

7 - توزيع كل من $\hat{\beta}_1$ و $\hat{\beta}_2$ مستقل عن $\hat{\sigma}^2$. وفي الفصل التالي ، سوف نقدم تفصيلاً لذلك .

8 - التقديران $\hat{\beta}_1$ و $\hat{\beta}_2$ لكل منهما أقل تباين في فئة التقديرات غير المتحيزة لكل منهما سواء كانت تقديرات خطية أو غير خطية . وهذه النتيجة ترجع إلى عالم الإحصاء Ros . وتعتبر هذه النتيجة ذات أهمية ، وذلك لأن نظرية ماركوف غير مقيدة بفئة التقديرات الخطية فقط(4) . لذلك تعتبر تقديرات المربعات الصغرى أفضل تقديرات غير متحيزة best unbiased estimators (BUE) لها أقل تباين .

وهنا يتضح أن افتراض التوزيع الطبيعي يمكننا من اشتقاق توزيعات المعاينة sampling distributions لكل من $\hat{\beta}_1$ و $\hat{\beta}_2$ (لكل منهما يتبع التوزيع الطبيعي) كذلك التقدير $\hat{\sigma}^2$ يتبع توزيع X^2 .

وفي الفصل التالي ، سوف نتناول تكوين فترات الثقة establishing confidence intervals وإجراء اختبارات الفروض testing hypotheses . ونلاحظ أن افتراض التوزيع الطبيعي للمتغير u_i بحيث :

$$u_i \sim N(0, \sigma^2)$$

يؤدي إلى أن Y (حيث Y_i دالة خطية في u_i) تتبع التوزيع الطبيعي أيضاً بحيث :

$$E(Yi) = \beta^1 + \beta^2 X_i \tag{7.3.4}$$

$$\text{var}(Y_i) = \sigma^2 \tag{8.3.4}$$

أي أن :

$$Y_i \sim N(\beta_1 + \beta_2 X_i,\ \sigma_2) \tag{9.3.4}$$

(4) C. R. Rao, Lnear Statistical Inference and Its Applications, John Wiley & Sons, New York, 1965, p. 258.

4.4 طريقة الإمكان الأعظم :
THE METHOD OF MAXIMUM LIKELIHOOD (ML)

طريقة الإمكان الأعظم طريقة للحصول على تقديرات ذات خصائص نظرية theortical properties أقوى من خصائص تقديرات المربعات الصغرى . وفي ملحق A4 في هذا الفصل نناقش بالتفصيل طريقة الإمكان الأعظم ، وتحت افتراض التوزيع الطبيعي للمتغير u_i نجد أن تقديرات المعلمات β_1 و β_2 باستخدام المربعات الصغرى متطابقة identical مع تقديرات الإمكان الأعظم . أما تقدير الإمكان الأعظم للمعلمة σ^2 حيث تساوي $\sum \hat{u}_i^2 / n$ تقدير متحيز biased في حين أن تقدير المربعات الصغرى للمعلمة σ^2 يساوي $\left(\sum \hat{u}_i^2 / (n-2)\right)$ غير متحيز unbiased ورغم اختلاف تقدير الإمكان الأكبر والمربعات الصغرى للمعلمة σ^2 بالنسبة لحجم العينات الصغيرة ولكن عندما $n \to \infty$ نجد أن التقديرين متساويان وغير متحيزين ، وتسمى هذه الخاصية بخاصية التقارب asymptotically .

وتعتبر طريقة المربعات الصغرى مع إضافة افتراض التوزيع الطبيعي للمتغير u_iأداة مهمة في تناول اختبارات الفروض ، وأقل تعقيداً من استخدام طريقة الإمكان الأكبر .

5.4 الملخص والاستنتاجات : SUMMARY AND CONCLUSION

1 – هذا الفصل ناقش نموذج الانحدار الخطي الطبيعي التقليدي classical normal linear regression model (CNLRM).

2 – هذا النموذج يختلف عن نموذج الانحدار الخطي التقليدي classical linear regression model (CLRM) في افتراض أن المتغير العشوائي u_i يتبع التوزيع الطبيعي . بالنسبة للنموذج (CLRM) لم يفترض افتراض التوزيع الطبيعي للمتغير العشوائي u_i ولكنه يفترض أن توقع u_i يساوي صفراً ، وتباينه مقدار ثابت محدد finite constant .

3 – الافتراض النظري للتوزيع الطبيعي يرجع إلى نظرية النزعة المركزية central limit theorem .

4 – وبدون افتراض التوزيع الطبيعي ، مع الأخذ في الاعتبار باقي الفروض الأخرى السابق مناقشتها في الفصل الثالث ، توضح نظرية جاوس ماركوف أن تقديرات المربعات الصغرى أفضل تقديرات خطية غير متحيزة BLUE .

5 – ومع إضافة افتراض التوزيع الطبيعي للمتغير u_i ، فإن تقديرات المربعات الصغرى تصبح أفضل تقديرات غير متحيزة best unbiased estimators (BUE) ذات توزيعات احتمالية معروفة . حيث تتبع كل من $\hat{\beta}_1$ و $\hat{\beta}_2$ التوزيعات الطبيعية أيضاً ، كذلك يتبع $\hat{\sigma}^2$ توزيع X^2 .

6 – في الفصلين 5 ، 8 سوف نوضح أهمية ماسبق ذكره أعلاه في إجراء الاستدلال الإحصائي حول معلمات المجتمع .

7 – تعتبر طريقة الإمكان الأكبر طريقة بديلة لطريقة المربعات الصغرى للحصول على تقديرات .

8 – في حالة افتراض التوزيع الطبيعي للمتغير العشوائي u_i ، تتساوى تقديرات المربعات الصغرى مع تقديرات الإمكان للمعلمتين $\hat{\beta}_1$ و $\hat{\beta}_2$ بينما تقدير المربعات الصغرى والإمكان الأعظم للمعلمة σ^2 مختلفان . ولكن كلما زاد حجم العينة يقترب التقديران من بعضهما .

9 – تعتبر طريقة الإمكان الأكبر ML طريقة عامة بالنسبة للعينات ذات الحجم الكبير large- sample method . كذلك تستخدم طريقة المربعات الصغرى في حالة النماذج غير الخطية في المعلمات nonlinear in the parameters ، ولمعرفة تفصيلات أكبر يرجع إلى الفصل 14 .

10 – في هذا المرجع ، سوف نعتمد على طريقة المربعات الصغرى في الحصول على التقديرات للأسباب العملية practical reasons التالية :

(أ) يعتبر استخدام المربعات الصغرى للحصول على تقديرات أسهل عملياً من استخدام طريقة الإمكان الأعظم .

(ب) تقديرات المربعات الصغرى والإمكان الأعظم متطابقان بالنسبة للمعلمات β_1 و β_2 .

(جـ) تقدير المعلمة σ^2 باستخدام طريقة المربعات الصغرى ، لايختلف كثيراً عن تقدير الإمكان في حالة العينات ذات الحجم الكبير .

ملحق A4 : APPENDIX

1.A4 تقديرات المربعات الصغرى لنموذج انحدار متغيرين :

Maximum likelihood estimation of two variable regression model

إذا اعتبرنا النموذج التالي $Y_i = \beta_1 + \beta_2 X_i + u_i$ فإذا كانت المتغيرات العشوائية u_i متغيرات عشوائية مستقلة لكل منها التوزيع الطبيعي (المعتاد) ، فإن المتغيرات Y_i متغيرات عشوائية مستقلة يتبع كل منها التوزيع الطبيعي (المعتاد) بتوقع يساوي $\beta_1 + \beta_2 X_i$ وتباين σ^2 (انظر المعادلة (9.3.4)) . وكنتيجة لذلك ، فإن دالة كثافة الاحتمال المشتركة the joint probability density function العشوائية $Y_1, Y_2, \dots, Y_n$ بشرط التوقع أو التباين السابقين يرمز لها بالرمز :

$$f(Y_1, Y_2, \dots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2)$$

وبما أن $Y_1, Y_2, \dots, Y_n$ متغيرات مستقلة ، وبالتالي فإن دالة كثافة الاحتمال المشتركة تصبح حاصل ضرب دوال كثافة الاحتمال للمتغيرات على النحو التالي :

$$\begin{aligned} & f(Y_1, Y_2, \dots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \\ & = f(Y_1 \mid \beta_1 + \beta_2 X_i, \sigma^2) f(Y_2 \mid \beta_1 + \beta_2 X_i, \sigma^2) \cdots f(Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) \end{aligned} \tag{1}$$

حيث

$$f(Y_i) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{2}$$

وتعتبر دالة كثافة الاحتمال $f(Y_i)$ دالة كثافة احتمال التوزيع الطبيعي (المعتاد) بتوقع $(\beta_1 + \beta_2 X_i)$ وتباين σ^2 . (حيث ترمز exp إلى الأساس الطبيعي e) . وبالتعويض في (1) بـ (2) نجد أن :

$$f(Y_i, Y_2, \dots, Y_n \mid \beta_1 + \beta_2 X_i, \sigma^2) = \frac{1}{\sigma^n\left(\sqrt{2\pi}\right)^n} \exp\left\{-\frac{1}{2}\sum \frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{3}$$

وعندما تكون قيم $Y_1, Y_2, ..., Y_n$ معلومة ، وقيم المعلمات $\beta_1, \beta_2, \sigma^2$ غير معلومة ، فإن الدالة (3) تسمى دالة الإمكان likelihood function ويرمز لها بالرمز LF $(\sigma^2, \beta_2, \beta_1)$ ، وتكتب على النحو التالي[1] :

$$\mathrm{LF}(\beta_1, \beta_2, \sigma^2) = \frac{1}{\sigma^n(\sqrt{2\pi})^n} \exp\left\{-\frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2}\right\} \tag{4}$$

وطريقة الإمكان الأكبر the method of maximum likelihood يمكن باستخدامها الحصول على تقديرات للمعلمات $\sigma^2, \beta_2, \beta_1$ ، هذه التقديرات التي تمكن من الحصول على أكبر احتمال ممكن بشرط القيم $Y_1, Y_2, ..., Y_n$ ويمكن الحصول على قيم تقديرات للمعلمات $\sigma^2, \beta_2, \beta_1$ بإيجاد النهاية العظمى للدالة (4) على النحو التالي ، ولتبسيط إيجاد المشتقات نوجد ln LF[2] أولاً :

$$\begin{aligned} \ln \mathrm{LF} &= -n\ln\sigma - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \\ &= -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_i)^2}{\sigma^2} \end{aligned} \tag{5}$$

وبإيجاد مشتقات الجزئية للدالة (5) بالنسبة $\sigma^2, \beta_2, \beta_1$ نحصل على

$$\frac{\partial \ln \mathrm{LF}}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)(-1) \tag{6}$$

$$\frac{\partial \ln \mathrm{LF}}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_i)(-X_i) \tag{7}$$

$$\frac{\partial \ln \mathrm{LF}}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(Y_i - \beta_1 - \beta_2 X_i)^2 \tag{8}$$

وبمساواة المعادلات (6)-(8) بالصفر نحصل على المعادلات التالية (بمساواة المعادلات بالصفر نحصل على الشرط الضروري لكي تكون الدالة LF وبالتالي ln LF نهاية عظمى) وبافتراض أن $\tilde{\sigma}^2, \tilde{\beta}_2, \tilde{\beta}_1$ تشير إلى تقديرات الإمكان بالتالي ، فإن[3] :

---

(1) Of course, if β_1, β_2, and σ^2 are known but the Y_t are not known, (4) represents the joint probability density function–the probability of jointly observing the Y_i.

(2) Since a log function is a monotonic function, ln LF will attain its maximum value at the same point as LF.

(3) We use ~ (tilde) for ML estimators and ^ (cap or hat) for OLS estimators.

$$\frac{1}{\tilde{\sigma}^2}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i) = 0 \quad (9)$$

$$\frac{1}{\tilde{\sigma}^2}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)X_i = 0 \quad (10)$$

$$-\frac{n}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 = 0 \quad (11)$$

وبتبسيط المعادلتين (9) ، (10) نحصل على :

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2 \sum X_i \quad (12)$$

$$\sum Y_i X_i = \tilde{\beta}_1 \sum X_i + \tilde{\beta}_2 \sum X_i^2 \quad (13)$$

ونلاحظ أن المعادلتين (12) ، (13) نفس المعادلات الطبيعية normal equations التي تم الحصول عليها باستخدام نظرية المربعات الصغرى the least squares theory في المعادلتين (4.1.3) ، (5.1.3) .

وبالتالي ، فإن تقديرات الإمكان $\tilde{\beta}_2, \tilde{\beta}_1$ ML هي نفسها تقديرات المربعات الصغرى $\tilde{\beta}_1$ و$\tilde{\beta}_2$ في المعادلتين (6.1.3) ، (7.1.3) .

وبالتعويض بتقديرات الإمكان ML (=OLS) في (11) يمكن الحصول على تقدير ML $\tilde{\sigma}^2$ على النحو التالي :

$$\begin{aligned} \tilde{\sigma}^2 &= \frac{1}{n}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_i)^2 \\ &= \frac{1}{n}\sum(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \quad (14) \\ &= \frac{1}{n}\sum \hat{u}_i^2 \end{aligned}$$

من العلاقة (14) يتضح أن التقدير$\tilde{\sigma}^2$ باستخدام ML يختلف عن$\tilde{\sigma}^2$ باستخدام المربعات الصغرى ، حيث :

$$\hat{\sigma}^2 = [1/(n-2)]\sum \hat{u}_i^2$$

حيث $\tilde{\sigma}^2$ تقدير غير متحيز unbiased للمعلمة σ^2 ، انظر ملحق (5A3) الفصل (3) . هكذا فإن تقديرات الإمكان للمعلمة σ^2 تقدير متحيز biased . ويمكن توضيح أن التقدير$\tilde{\sigma}^2$ تقدير متحيز على النحو التالي :

$$E(\tilde{\sigma}^2) = \frac{1}{n}E\left(\sum \hat{u}_i^2\right)$$
$$= \left(\frac{n-2}{n}\right)\sigma^2 \qquad (15)$$
$$= \sigma^2 - \frac{2}{n}\sigma^2$$

ويتضح أن $\tilde{\sigma}^2$ تقدير متحيز (بمعنى أنه تقدير أقل من القيمة الحقيقية σ^2 في العينات صغيرة الحجم . ولكن يلاحظ أن الحد الثاني في (15) يؤول إلى الصفر عندما يزداد حجم العينة ، أي عندما $n \to \infty$ ، أو بعبارة أخرى $E(\tilde{\sigma}^2) = \sigma^2$ عندما $n \to \infty$ ، كذلك يمكن إثبات أن $\tilde{\sigma}^2$ تقدير متسق consistent[4] .

2.A4 تقديرات الإمكان الأعظم للإنفاق على الغذاء في الهند :

Maximum likelihood estimation of food expenditure in India

بالرجوع إلى مثال 3-2 ونموذج الانحدار (2.7.3) الذي يوضح العلاقة بين الإنفاق على الغذاء كدالة في الإنفاق الكلي total expenditure لـ 55 أسرة بالريف في الهند . وتحت افتراض التوزيع الطبيعي لـ u_i تم الحصول على تقديرات المربعات الصغرى (المطابقة لتقديرات الإمكان الأعظم) للمعلمات β_1 و β_2 ، حيث تم الحصول على تقديرات ML على النحو التالي :

$$\tilde{\beta}_1 = \hat{\beta}_1 = 94.2087, \ \tilde{\beta}_2 = \hat{\beta}_2 = 0.4386$$

كذلك تقديرات المربعات الصغرى للمعلمة σ^2 على النحو :

$$\hat{\sigma}^2 = 4469.6913$$

كذلك تقديرات σ^2 باستخدام ML تصبح على النحو :

$$\tilde{\sigma}^2 = 4407.1563$$

---

(4) See App. A for a general discussion of the properties of the maximum likelihood estimators as well as for the distinction between asymptotic unbiasedness and consistency. Roughly speaking, in asymptotic unbiasedness we try to find out the lim $E(\tilde{\sigma}_n^2)$ as n tends to infinity, where n is the sample size on which the estimator is based, whereas in consistency we try to find out how $\tilde{\sigma}_n^2$ behaves as n increases indefinitely. Notice that the unbiasedness property is a repeated sampling property of an estimator based on a sample of given size, whereas in consistency we are concerned with the behavior of an estimator as the sample size increases indefinitely.

حيث نجد أن تقدير الإمكان الأعظم للمعلمة σ^2 أقل من تقدير المربعات الصغرى لنفس المعلمة . مع ملاحظة أن العينات الصغيرة ، تقدير الإمكان الأعظم للمعلمة σ^2 يكون متحيزًا وأقل من قيمة المعلمة σ^2 .

تمرينات على ملحق A4 : Appendix exercises

1.4 "إذا كان متغيران عشوائيان مستقلان إحصائياً ، فإن معامل الارتباط بينهما يساوي صفرًا . ولكن العكس ليس بالضرورة صحيح ، بمعنى إذا كان معامل الارتباط يساوي صفرًا ، فهذا لايعني بالضرورة استقلالهما إحصائياً . ومع ذلك لو متغيرين يتبع كل منهما التوزيع الطبيعي (المعتاد) ومعامل الارتباط بينهما يساوي صفرًا ، فهذا يؤول إلى أن المتغيرين مستقلان إحصائياً" . اثبت هذه العبارة بالنسبة لدالة كثافة الاحتمال المشتركة لمتغيرين Y_1, Y_2 حيث يتبع كل منهما التوزيع المعتاد (الطبيعي) (دالة كثافة الاحتمال المشتركة التالية تسمى بدالة كثافة الاحتمال المشتركة للتوزيع الطبيعي الثنائي the bivariate normal probability density function) :

$$f(Y_1, Y_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\right.$$

$$\left. \times \left[\left(\frac{Y_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{(Y_1-\mu_1)(Y_2-\mu_2)}{\sigma_1\sigma_2} + \left(\frac{Y_2-\mu_2}{\sigma^2}\right)^2\right]\right\}$$

حيث :

توقع المتغير Y_1 $= \mu_1$

توقع المتغير Y_2 $= \mu_2$

الانحراف المعياري للمتغير Y_2 $= \sigma_1$

الانحراف المعياري للمتغير Y_2 $= \sigma_2$

معامل الارتباط بين المتغيرين Y_1 و Y_2 $= \rho$

2.4 طبق الشروط من الترتيب الثاني للحصول على النهاية العظمى لدالة الإمكان (بمعنى إجراء اختبار المشتقات من الترتيب الثاني) ، لتوضيح أن تقديرات الإمكان الأعظم للمعلمات $\sigma^2, \beta_2, \beta_1$ التي يتم الحصول عليها بحل المعادلات (9)-(11) ، تحقق النهاية العظمي لدالة الإمكان في (4) .

3.4 اعتبر المتغير العشوائي X حيث X يتبع التوزيع الأسي exponential distribution ، بدالة كثافة الاحتمال (PDF) التالية :

$$f(X) = (1/\theta)e^{-X/\theta} \qquad X > 0 \ , \ \ 070$$

$= 0$ فيما عدا ذلك

استخدم طريقة ML لتوضيح أن تقدير ML للمعلمة θ حيث :

$$\hat{\theta} = \sum X_i / n$$

حيث n حجم العينة . اثبت أن θ تساوي متوسط العينة $\bar{X}$.

الفصل الخامس

انحدار متغيرين: تقدير الفترة واختبارات الفروض

TWO - VARIABLE REGRESSION: INTERVAL ESTIMATION AND HYPOTHESIS TESTING

في هذا الفصل ، بجانب تناول تقديرات الفترة ، سوف نتناول أيضاً اختبارات الفروض المبنية على البيانات[1].

وكما ذكرنا في الفصل السابق ، أن تقديرات واختبارات الفروض تمثل الفرعين الرئيسين لعلم الإحصاء. وتنقسم نظرية التقدير the theory of estimation إلى جزءين هما : تقدير النقطة، وتقدير الفترة point estimation and interval estimation. وقد تناولنا في الفصلين السابقين ، تقديرات النقطة باستخدام طريقتي المربعات الصغرى OLS ، والإمكان الأعظم ML.

وفي هذا الفصل ، سوف نتناول تقديرات الفترة ، ثم نتناول اختبارات الفروض المرتبطة باختبارات الفترة.

1.5 المتطلبات الإحصائية : STATISTICAL PRERQUISTES

وقبل تناول تقديرات الفترة ، واختبارات الفروض ، فإن ذلك يتطلب أن يكون القارئ ملمًا بالمفاهيم الأساسية للاحتمالات والإحصاء.

بالإضافة إلى كورس أساسي في الإحصاء، ملحق A يمد القارئ بالأساسيات الإحصائية الضرورية للقارئ ، مثل بعض التعريفات والمفاهيم الأساسية مثل

(1) Stephen M. Stigler, "Testing Hypothesis or Fitting Models? Another Look at Mass Extinctions," in Matthew H. Nitecki and Antoni Hoffman, eds., Neutral Models in Biology, Oxford University Press, Oxford, 1987, p. 148.

الاحتمال probability، التوزيعات الاحتمالية probability distributions، أخطاء النوعين الأول والثاني type I and type II errors، مستوى المعنوية level of significance، قوة الاختبار الإحصائي power of a statistical test، فترة الثقة confidence interval . فمثل هذه المفاهيم والتعريفات ضرورية لفهم هذا الفصل .

2.5 تقدير الفترة.. بعض الأفكار الأساسية :
INTERVAL ESTIMATION: SOME BASIC IDEAS

لتحديد الأفكار الرئيسية، سوف نعتبر المثال الافتراضي للدخل والاستهلاك في الفصل 3. المعادلة (2.6.3) توضح أن تقدير الميل الحدي للاستهلاك $\hat{\beta}_2$ marginal propensity to consume (MPC) يساوي 0.5091، حيث يعتبر 0.5091 تقدير نقطة point estimate للمعلمة β_2. والسؤال الآن مامدى صلاحية هذا التقدير؟

وكما سبق أن أوضحنا في الفصل الثالث: أن قيمة التقدير $\hat{\beta}_2$ تختلف من عينة إلى عينة أخرى، ولكن القيمة المتوقعة للقيم المختلفة لـ $\hat{\beta}_2$ تساوي قيمة المعلمة β_2 ، حيث سبق أن أوضحنا أن $E(\hat{\beta}_2)=\beta_2$. وكما سبق أن ذكرنا أن صلاحية تقدير النقطة $\hat{\beta}_2$ يمكن أن تقاس بالخطأ المعياري standard error للمتغير $\hat{\beta}_2$.

والآن سوف نقوم بتحديد الفترة التي تحيط بتقدير النقطة ، بحيث تتضمن هذه الفترة القيمة الحقيقية للمعلمة β_2 بنسبة معينة (أو باحتمال معين) ، وتعتبر هذه الفكرة الأساسية وراء تقدير الفترة interval estimation .

ولتحديد أكثر، سوف نفترض أن المعلمة β_2 تقع داخل الفترة $\hat{\beta}_2-\delta,\ \hat{\beta}_2+\delta$ باحتمال $(1-\alpha)$ حيث δ , α قيم موجبة تقع بين الصفر والواحد. أو بعبارة أخرى

$$\Pr(\hat{\beta}_2-\delta \le \beta_2 \le \hat{\beta}_2+\delta) = 1-\alpha \qquad (1.2.5)$$

حيث تسمى الفترة $(\hat{\beta}_2-\delta,\ \hat{\beta}_2+\delta)$ بفترة الثقة confidence interval، ويسمى الاحتمال $(1-\alpha)$ بمعامل الثقة confidence coefficient، أويسمى الاحتمال (α) بمستوى المعنوية level of significance[2] وتسمى الحدود $(\hat{\beta}_2+\delta)$, $(\hat{\beta}_2-\delta)$ بحدود الثقة confidence limits (وتسمى أحياناً بالقيم الحرجة critical values)، حيث يمثل $(\hat{\beta}_2-\delta)$

(2) Also known as the probability of committing a Type I error. A Type I error consists in rejecting a true hypothesis whereas a Type II error consists in accepting a false hypothesis. (This topic is discussed more fully in App.A) The ymbol α is also known as the size of the (statistical) test.

بالحد الأدنى للثقة lower confidence limit كذلك $(\hat{\beta}_2 + \delta)$ بالحد الأعلى للثقة upper confidence limit والمعادلة (1.2.5) توضح تقدير الفترة interval estimator. وأنه من الأهمية معرفة الجوانب التالية :

1 - باستخدام المعادلة (1.2.5) يمكن تحديد حدود الثقة عند معامل ثقة $(1-\alpha)$ (أو بعبارة أخرى مستوى معنوية α).

2 - الفترة في المعادلة (1.2.5) تعتبر فترة عشوائية random interval ، بمعنى أن هذه الفترة تختلف من عينة لأخرى المسحوب منها $\hat{\beta}_2$. (لماذا؟).

3 - بما أن فترة الثقة عشوائية ، فإن متوسط هذه الفترات باحتمال كل منها يساوي $(1-\alpha)$ يعني أن باحتمال $(1-\alpha)$ تقع المعلمة β_2 داخل هذه الفترة (أي باحتمال $(1-\alpha)$).

4 - كما أوضحنا في 2، أن الفترة في (1.2.5) تكون عشوائية طالما أن $\hat{\beta}_2$ غير معلومة. ولكن عند اختيار عينة محددة وحساب $\hat{\beta}_2$ فإن الفترة في (1.2.5) تصبح فترة محددة وليست عشوائية.

فبالنسبة **للمثال الافتراضي للدخل والاستهلاك** ، إذا كانت $1-\alpha = 0.95$ أي $1-\alpha = \%95$ فإن فترة الثقة $(0.4268 \le \beta_2 \le 0.5914)$ كما سوف يوضح في المعادلة (9.2.5). ويتبقى أن نعرف كيف يمكن تحديد فترات الثقة؟ ومن المناقشة السابقة يمكن أن نتوقع أن لو التوزيعات الاحتمالية للمعاينة the sampling or probability distributions للتقديرات معلومة known يمكن إيجاد فترة الثقة كما في المعادلة (1.2.5). وكما ذكرنا في الفصل (4)، تحت فرض التوزيع الطبيعي للمتغير العشوائي u_i فإن تقديرات المربعات الصغرى $\hat{\beta}_1$ و$\hat{\beta}_2$ تتبع التوزيع الطبيعي أيضاً، كذلك تقدير المربعات الصغرى للتباين $\hat{\sigma}^2$ متغير مرتبط بتوزيع مربع كا (X^2) حيث يتطلب تحديد فترة الثقة معرفة التوزيع الاحتمالي لـ $\hat{\sigma}^2$ و$\hat{\beta}_2$ و$\hat{\beta}_1$.

3.5 فترات الثقة لمعلمات الانحدار β_1 و β_2 : CONFIDENCE INTERVALS FOR REGRESSION COEFFICIENTS β_1 AND β_2

من الفصل الرابع (3-4) يتضح أنه تحت افتراض التوزيع الطبيعي للمتغير u_i فإن كلاً من المعلمات $\hat{\beta}_1$ و $\hat{\beta}_2$ يتبع التوزيع المعتاد (الطبيعي) أيضاً ، وبالتالي فإن :

$$Z = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)}$$

$$= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\sigma} \qquad (1.3.5)$$

ومن المعادلة (6.3.4) نجد أن المتغير Z يتبع التوزيع المعتاد القياسي (أي توزيع معتاد بتوقع يساوي صفراً وتباين يساوي واحداً)، إذا كان التباين σ^2 معلومة. ومن خصائص التوزيع المعتاد (الطبيعي) بتوقع يساوي μ وتباين يساوي σ^2 أي أن المساحة المحصورة بين منحنى دالة كثافة الاحتمال للمتغير المعتاد والمحور الأفقي بين $\mu \pm \sigma$ تساوي 68% كذلك المساحة المحصورة بين $\mu \pm 2\sigma$ تساوي 95%، كذلك بين الحدين $\mu \pm 2\sigma$ تساوي 99.7% من المساحة الكلية أسفل المنحنى.

ولكن عادة σ^2 تكون عملياً غير معلومة ، ويستخدم التقدير غير المتحيز $\hat{\sigma}^2$ تقديراً لها. فإذا تم إحلال $\hat{\sigma}$ بدلاً من σ في (1.3.5) فإنه يمكن كتابة العلاقة على النحو التالي :

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} = \frac{\text{estimator} - \text{parameter}}{\text{estimated standard error of estimator}} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \qquad (2.3.5)$$

حيث ترمز $\text{se}(\hat{\beta}_2)$ إلى تقدير الخطأ المعياري estimated standard error. ويتضح أن المتغير t (انظر ملحق 2A5 الفصل 5) يتبع توزيع أستيودنت ، أي توزيع t بدرجات حرية $n-2$ [لاحظ الفرق بين (1.3.5)، (2.3.5)]. لذلك يستخدم التوزيع t بدلاً من التوزيع المعتاد لتقدير فترة الثقة للمعلمة B_2 على النحو التالي :

$$\Pr(-t_{\alpha/2} \le t \le t_{\alpha/2}) = 1 - \alpha \qquad (3.3.5)$$

حيث $t_{\alpha/2}$ تشير إلى حد الثقة عند مستوى معنوية يساوي α ، ويتم الحصول عليها من جدول توزيع t عند درجة حرية $(n-2)$ ومستوى معنوية α.

وبالتعويض في (2.3.5) في (3.3.5) نحصل على :

$$\Pr\left[-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} \le t_{\alpha/2}\right] = 1 - \alpha \qquad (4.3.5)$$

ويمكن إعادة صياغة (4.3.5) نجد أن :

$$\Pr[\hat{\beta}_2 - t_{\alpha/2}\,\text{se}(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{\alpha/2}\,\text{se}(\hat{\beta}_2)] = 1 - \alpha \qquad {}^{(3)}(5.3.5)$$

(3) Some authors prefer to write (5.3.5) with the df explicitly indicated. Thus, they would write

$$\Pr[\hat{\beta}_2 - t_{(n-2),\alpha/2}\,\text{se}(\hat{\beta}_2) \le \beta_2 \le \hat{\beta}_2 + t_{(n-2)\alpha/2}\,\text{se}(\hat{\beta}_2)] = 1 - \alpha$$

But for simplicity we will stick to our notation; the context clarifies the appropriate df involved.

والمعـادلة (5.3.5) توضح النسبـة المئـوية $(1-\alpha)$ 100 لفـتـرة الثـقـة confidence interval للمعلمة β_2 والتي تكتب وتقرأ على النحو :

بدرجة ثقة $(1-\alpha)$ 100 فترة الثقة للمعلمة β_2 على النحو التالي :

$$\hat{\beta}_2 \pm t_{\alpha/2}\, \text{se}(\hat{\beta}_2) \tag{6.3.5}$$

وبنفس الطريقة يمكن إثبات أنه باستخدام المعادلتين (1.3.4)، (2.3.4) على النحو :

$$\Pr[\hat{\beta}_1 - t_{\alpha/2}\, \text{se}(\hat{\beta}_1) \leq \beta_1 \leq \hat{\beta}_1 + t_{\alpha/2}\, \text{se}(\hat{\beta}_1)] = 1-\alpha \tag{7.3.5}$$

ويمكن إعادة كتابة المعادلة على النحو التالي :

بدرجة ثقة $(1-\alpha)\%$ 100 فترة الثقة للمعلمة β_1 على النحو :

$$\hat{\beta}_1 \pm t_{\alpha/2}\, \text{se}(\hat{\beta}_1) \tag{8.3.5}$$

ويلاحظ أن فترات الثقة التي تم اشتقاقها في العلاقات (8.3.5) و (6.3.5) تلعب دوراً بالغ الأهمية في الاستدلال. ويلاحظ أن اتساع فترة الثقة مرتبط بنسبة الخطأ المعياري للتقدير standard error of the estimator.

وفي العلاقتين (8.3.5)، (6.3.5) يتضح أنه كلما زاد الخطأ المعياري للتقدير أدى ذلك إلى زيادة اتساع فترة الثقة- لذا غالباً يحدد الخطأ المعياري للتقدير دقة التقدير the precision of the estimator.

وبالرجوع إلى المثال التوضيحي للدخل ، والاستهلاك في (الفصل 6.3) نجد أن :

$$\hat{\beta}_2 = 0.5091, \text{ se}(\hat{\beta}_2) = 0.0357$$

وعدد درجات الحرية (df) = 8

وإذا فرضنا أن 5% = α أي أن معامل الثقة confidence coefficient يساوي 95%، والقيمة الجدولية من جدول توزيع t عند درجات الحرية 8 $t_{\alpha/2}$ حيث :

$$t_{\alpha/2} = t_{0.025} = 2.306$$

وبالتعويض بهذه القيم في (5.3.5) نجد أن :

$$0.4268 \leq \beta_2 \leq 0.5914 \tag{9.3.5}$$

كذلك بالتعويض في (6.3.5) نجد أن :

$$0.5091 \pm 2.306(0.0357)$$

أو

$$0.5091 \pm 0.0823 \qquad (10.3.5)$$

وتفسير لفترة الثقة يكون على النحو التالي :

عند درجة الثقة (معامل الثقة) confidence coefficient تساوي 95%- تعني في المدى الطويل كل 100 حالة يوجد 95 حالة تكون قيمة β_2 تقع في الفترة (0.4268 , 0.5914).

فترة الثقة للمعلمة β_1 : Confidence Interval for

في المعادلة (7.3.5) يمكن إثبات أن فترة الثقة للمعلمة β_1 عند معامل الثقة (درجة الثقة) 95% في مثال الدخل والاستهلاك على النحو التالي :

$$9.6643 \le \beta_1 \le 39.2448 \qquad (11.3.5)$$

ويمكن إعادة كتابة العلاقة (8.3.5) على النحو :

$$24.4545 \pm 2.306(6.4138)$$
$$\rightarrow$$
$$24.4545 \pm 14.7902 \qquad (12.3.5)$$

ويمكن تفسير الفترة للمعلمة β_1 على النحو التالي : عند درجة الثقة 95% تكون فترة الثقة التي تقع فيها المعلمة β_1 كما في العلاقة (1.3.5) وهذا يعني في المدى الطويل كل 100 حالة يكون فيها 95 حالة تكون قيمة β_1 تقع داخل الفترة في (11.3.5).

فترة الثقة لكل من β_1 و β_2 آنياً :

Confidence intervals for β_1 and β_2 simultaneously

وفي بعض الحالات يكون من الأهمية الحصول على فترة ثقة joint confidence interval لكل من β_1 و β_2 معاً عند معامل ثقة $(1-\alpha)$. فترات الثقة من هذا النوع تتطلب الرجوع إلى مراجع[4] متخصصة. وفي الأبواب 8 ، 9 ، 10 في هذا المرجع ، سوف نتناول باختصار شديد هذا النوع من فترات الثقة.

(4) For an accessible discussion, see John Neter, William Wasserman, and Michael H. Kutner, Applied Linear Regression Models, Richard D. Irwin, Homewood, Ill., 1983, Chap. 5.

4.5 فترة الثقة للتباين σ^2 : CONFIDENCE INTERVAL FOR σ^2

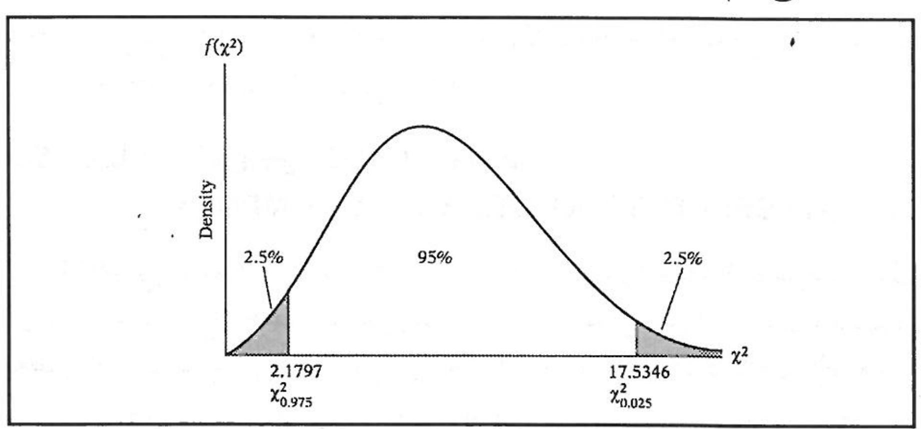
وكما أشرنا سابقاً في الفصل (3.4) ، وتحت فرض التوزيع الطبيعي (المعتاد) للمتغير u_i فإن المتغير X^2 حيث :

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \tag{1.4.5}$$

يتبع توزيع مربع كا^2 (أي توزيع X^2) بدرجات حرية $(n-2)$[5]. وبالتالي يمكن استخدام توزيع كا^2 (X^2) لإيجاد فترة ثقة لـ σ^2 على النحو التالي :

$$\Pr\left(\chi^2_{1-\alpha/2} \le \chi^2 \le \chi^2_{\alpha/2}\right) = 1-\alpha \tag{2.4.5}$$

حيث X^2 قيمة تقع داخل الفترة $[X^2_{1-a/2}, X^2_{a/2}]$ في المعادلة (2.4.5) حيث تم إيجاد قيم $X^2_{1-a/2}$, $X^2_{a/2}$ من جدول توزيع X^2 بدرجات حرية $(n-2)$. وشكل (1-5) يوضح قيم X^2_{1-a} , $X^2_{a/2}$ عند $100(1-\alpha)\%$



شكل (1.5) القيم التي تقع داخل لفترة

وبالتعويض في (2.4.5) بـ (1.4.5) نحصل على :

$$\Pr\left[(n-2)\frac{\hat{\sigma}^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le (n-2)\frac{\hat{\sigma}^2}{\chi^2_{1-\alpha/2}}\right] = 1-\alpha \tag{3.4.5}$$

والمعادلة (3.4.5) توضح أن المعلمة σ^2 تقع داخل الفترة $\left[(n-2)\frac{\hat{\sigma}^2}{X^2_{a/2}}, (n-2)\frac{\hat{\sigma}^2}{X^2_{1-a/2}}\right]$ باحتمال $(1-\alpha)$.

(5) For proof, see Robert V. Hogg and Allen T. Craig, Introduction to Mathematical Statistics, 2d ed., Macmillan, New York, 1965. p. 144.

ولتوضيح ذلك من خلال المثال في الفصل (6.3) نجد أن $\hat{\sigma}^2 = 42.1591$ بدرجات حرية (df) تساوي 8. وعندما $\alpha = 0.05$ أي $\alpha = \%5$ من جداول توزيع X^2 وعند درجات حرية تساوي 8 نجد أن $X^2_{0.025} = 17.5346,\ X^2_{0.975} = 2.1797$. وهذه القيم توضح احتمال أن قيمة X^2 ، تزيد عن 17.5346 تساوي %2.5، كذلك احتمال أن قيمة X^2، واحتمال أن تزيد عن 2.1797 يساوي %97.5. لذلك الفترة بين هاتين القيمتين (17.5340 , 2.1797) بدرجة ثقة %95 كما هو موضح بشكل (1.5) (لاحظ التواء توزيع مربع كا).

وبالتعويض ببيانات المثال في (3.4.5) يمكن إثبات أن فترة الثقة للمعلمة σ^2 عند درجة ثقة %95 على النحو التالي :

$$19.2347 \le \sigma^2 \le 154.7336 \qquad (4\text{-}4\text{-}5)$$

وتفسير هذه الفترة هو أنه عند درجة ثقة (معامل ثقة) %95، قيمة المعلمة σ^2 تقع داخل الفترة في (4.4.5) أو بعبارة أخرى إذا كان لدينا 100 حالة فإنه في 95 حالة منها تكون قيمة σ^2 تقع داخل الحدود في (4.4.5).

5.5 اختبارات الفروض (تعليقات عامة) :
HYPOTHESIS TESTING: GENERAL COMMENTS

تناولنا في الفصول السابقة ، تقديرات النقطة ، وتقديرات الفترة. وسوف نتناول فيما يلي ، موضوع اختبارات الفروض. في هذا الفصل ، سوف نتناول باختصار بعض الجوانب العامة لهذا الموضوع، وملحق A يوضح بعض التفصيلات الإضافية.

مشكلة اختبارات الفروض الإحصائية statistical hypothesis testing ممكن أن تحدد ببساطة على النحو التالي : هل المشاهدة المعطاة a given observation أو البيانات (النتيجة) ملائمة finding compatible مع فرض محدد ما أم لا؟ وكلمة ملائمة "compatible" تستخدم هنا بمعنى الاقتراب الكافي sufficiently من القيمة الافتراضية hypothesized value لكي نرفض الفرض المحدد the stated hypothesis.

هكذا، لو نظرية ما أو خبرة سابقة أدت إلى اعتقاد أن القيمة الحقيقية للمعلمة β_2 في مثال الدخل والاستهلاك تكون قيمة وحيدة، هل القيمة المشاهدة $\hat{\beta}_2 = 0.5091$ التي تم الحصول عليها من العينة في جدول (2.3) متسقة مع الفرض المحدد؟ ولو كان كذلك لا تستطيع رفض الفروض، فيما عدا ذلك يمكن رفضه.

وبلغة الإحصاء، الفرض المحدد stated hypothesis ويرمز له بالرمز H_0. والفرض العدمي عادة يختبر ضد فرض بديلة alternative hypothesis ويرمز له بالرمز H_1، فعلى سبيل المثال(6) :

$$H_0 : \beta_2 = 1.5 \quad , \quad H_1 : \beta_2 \neq 1.5$$

ونظرية اختبارات الفروض تهتم بتطوير قواعد أو إجراءات لتقرير رفض أو رفض الفرض العدمي.

ويوجد أسلوبان متنافيان ومتكاملان two mutually complementary approaches لإعداد هذه القواعد، بمعنى فترة الثقة confidence interval ، واختبارات المعنوية test of significance . كل أسلوب من هذين الأسلوبين يرتبط بتقدير (إحصاء) يتغير تحت اعتبارات تتمثل في التوزيع الاحتمالي للتقدير واختبارات الفرض، بحيث يتم وضع صياغة أو صياغات statements حول قيمة (أو قيم) المعلمة (أو المعلمات) لهذا التوزيع (أو التوزيعات). فعلى سبيل المثال، تحت افتراض التوزيع المعتاد u_i نجد أن التقدير $\hat{\beta}_2$ يتبع التوزيع المعتاد أيضاً بتوقع β_2 وتباين موضح في العلاقة (4.3.5). فإذا افترضنا أن $\beta_2 = 1$.

وفي هذا الكتاب، سوف نتناول اختبارات الفروض لمعلمات التوزيعات الاحتمالية، وبصفة خاصة بالنسبة لمعلمات التوزيعات الاحتمالية، وبصفة خاصة بالنسبة لمعلمات التوزيع المعتاد (الطبيعي).

6.5 اختبارات الفروض.. أسلوب فترات الثقة : HYPOTHESIS TESTING: THE CONFIDENCE INTERVAL APPROACH

اختبار الذيلين : Two- sided or two- tail test

لتوضيح أسلوب فترات الثقة، سوف نوضح ذلك من خلال مثال الدخل، والاستهلاك على النحو التالي. سبق توضيح أن تقدير المعدل الهامشي للاستهلاك (MPC) $\hat{\beta}_2$ يساوي 0.5091 . والآن سوف نعتبر الفرض العدمي والفرض البديل على النحو التالي :

---

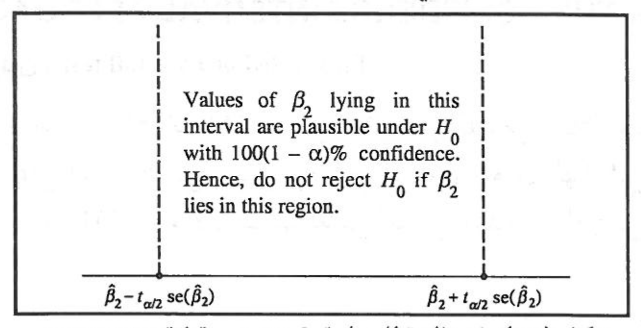
(6) A statistical hypothesis is called a simple hypothesis if is specifies the precise value(s) of the parameter(s) of a probability density function; otherwise, it is called a composite hypothesis. For example, in the normal pdf $(1/\sigma\sqrt{2\pi})\exp\{-\frac{1}{2}[(X-\mu)/\sigma]^2\}$, if we assert that H_1: μ=15 and σ=2, it is a simple hypothesis; but if H_1: μ=15 and σ > 15, it is a composite hypothesis, because the standard deviation does not have a specific value.

$$H_0 : \beta_2 = 0.3$$
$$H_1 : \beta_2 \neq 0.3$$

والفرض العدمي H_0 يعني أن القيمة الحقيقية لـ β_2 تساوي 0.3 ، والفرض البديل H_1 يعني أن القيمة الحقيقية لـ β_2 لاتساوي 0.3 (أي أكبر من 0.3 أو أقل من 0.3). وفي حالة عندما يكون الفرض البديل في صيغة ≠ فإن الاختبار في هذه الحالة يسمى اختبار الذيلين (وسوف نوضح ذلك فيما بعد) وفي هذه الحالة يسمى الفرض البديل H_1 فرض مركب composite وهنا نطرح سؤالاً : هل القيمة $\hat{\beta}_2$ متوافقة compatible مع الفرض العدمي H_0؟ وللإجابة على هذا السؤال ، دعنا نرجع إلى فترة الثقة في العلاقة (9.3.5). حيث وجدنا أنه في المدى الطويل أن β_2 تقع داخل الفترة (0.4268 , 0.5919) ، وذلك بدرجة ثقة 95% . لذلك إذا كانت قيمة β_2 في الفرض العدمي H_0 تقع داخل فترة الثقة بدرجة ثقة $100(1-\alpha)\%$ فهذا يعني عدم رفض الفرض العدمي H_0، أما إذا وقعت قيمة β_2 في الفرض العدمي H_0 خارج فترة الثقة ، فإننا نحن نرفض الفرض العدمي H_0[7]. وهذا المدى لفترة الثقة موضح في شكل (2.5).

قاعدة القرار : عند درجة ثقة $100(1-\alpha)\%$ فترة الثقة للمعلمة β_2. إذا كانت β_2 في الفرض العدمي H_0 داخل فترة الثقة ، فإنه لايمكن رفض الفرض العدمي. وفي حالة وقوع قيمة β_2 في الفرض العدمي H_0 خارج فترة الثقة ، فإنه يتم رفض الفرض العدمي H_0.

ووفقاً لهذه القاعدة القرارية ، فإن الفرض العدمي $H_0 = 0.3$ واضح أنه يقع خارج فترة الثقة عند درجة الثقة 95% في المثال محل الدراسة ، لذا نرفض الفرض العدمي



شكل (2.5) فترة الثقة للمعلمة β_2 عند درجة ثقة $100(1-\alpha)\%$

(7) Always bear in mind that there is a 100a percent chance that the confidence interval does not contain β_2 under H0 even though the hypothesis is correct. In short, there is a 100a percent chance of committing a **type I error**. Thus, if $a = 0.05$, there is a 5 percent chance that we could reject the null hypothesis even though it is true.

وفي الإحصاء عندما نرفض الفرض العدمي ، فإنه يقال إن الرقم في الفرض العدمي معنوي إحصائياً statically significant ويقال إن الرقم في الفرض العدمي غير معنوي إحصائياً not statistically significant في حالة عدم رفض الفرض العدمي .

اختبار الذيل الواحد : One- Sided or One Tail Test

وأحياناً يكون لدينا توقع نظري theoretical expectation أو خبرة سابقة تمكننا من وضع الفرض البديل في اتجاه واحد فقط . فمثلاً في مثال الدخل والاستهلاك يمكن كتابة الفرض العدمي ، والفرض البديل على النحو التالي :

$$H_0 : \beta_2 \leq 0.3 \,, \quad H_1 : \beta_2 > 0.3$$

وربما من النظرية الاقتصادية يمكن اقتراح أن معامل الاستهلاك الحدي أكبر من 0.3. ويمكن إجراء هذا الاختبار باستخدام العلاقة (5.3.5)[8].

7.5 اختبارات الفروض.. أسلوب اختبار المعنوية :
HYPOTHESIS TESTING: THE TEST OF SIGNIFICANCE APPROACH

اختبار المعنوية لمعاملات الانحدار.. اختبار t :

Testing the significance of regression coefficients: the t test

ويوجد أسلوب آخر مكمل complementary approach لأسلوب فترة الثقة the confidence interval method يسمى بأسلوب اختبار المعنوية the test of significance approach تم تطويره كل من R.A.Fisher ، Neyman and parson[9] كل منهم على حدة .

ويمكن القول باختصار ، إن اختبار المعنوية هو إجراء procedure. باتباعه يمكن استخدام نتائج العينة لإثبات صحة أو خطأ الفرض العدمي a null hypothesis . والفكرة الأساسية لإجراء اختبار المعنوية تكمن في اختبار الإحصاء a test statistic (estimator) وتوزيع المعاينة sampling distribution للإحصاء (التقدير) محل الاعتبار في الفرض العدمي .

وقرار قبول أو رفض الفرض العدمي H_0 يبنى أساساً على قيمة إحصاء الاختبار test statistic الذي يتم الحصول عليه باستخدام بيانات العينة . وللتوضيح، سوف نعتبر

(8) If you want to use the confidence interval approach, construct a $(100-\alpha)\%$ one-sided or one-tail confidence interval for β_2. Why?

(9) Details may be found in E. L. Lehman, Testing Statistical Hypotheses, John Wiley & Sons, New York, 1959.

فرض التوزيع الطبيعي للخطأ العشوائي u_i، وتحت هذا الفرض نجد أن المتغير t حيث :

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)}$$

$$= \frac{(\hat{\beta}_2 - \beta_2)\sqrt{\sum x_i^2}}{\hat{\sigma}} \qquad (2.3.5)$$

حيث المتغير t متغير يتبع توزيع استيودنت (t) بدرجات حرية $(n-2)$. وبالتالي عند درجة ثقة $(1-\alpha)$ نجد أن :

$$\Pr\left[-t_{\alpha/2} \le \frac{\hat{\beta}_2 - \beta_2^*}{\text{se}\,(\hat{\beta}_2)} \le t_{\alpha/2}\right] = 1 - \alpha \qquad (1.7.5)$$

حيث β_2^* تشير إلى قيمة المعلمة β_2 تحت الفرض العدمي H_0. ومن جدول توزيع t عند درجات حرية $(n-2)$ ودرجة ثقة $1-\alpha$ يمكن الحصول على قيم حدود الثقة $\pm t_{\alpha/2}$، وتوجد جداول توزيع t بملحق D. ومن المعادلة (1.7.5) يمكن إعادة كتابتها على النحو التالي :

$$\Pr\left[\beta_2^* - t_{\alpha/2}\,\text{se}(\hat{\beta}_2) \le \hat{\beta}_2 \le \beta_2^* + t_{\alpha/2}\,\text{se}(\hat{\beta}_2)\right] = 1 - \alpha \qquad (2.7.5)$$

والمعادلة (2.7.5) تعطي فترة الثقة للمتغير $\hat{\beta}_2$ باحتمال $1-\alpha$ بشرط $\beta_2 = \beta_2^*$. وبلغة اختبارات الفروض وتسمى الفترة $\left[\beta_2^* - t_{\alpha/2}\,\text{se}(\hat{\beta}_2)\,,\ \beta_2^* + t_{\alpha/2}\,\text{se}(\hat{\beta}_2)\right]$ بفترة الثقة أو بمنطقة القبول the region of acceptance (للفرض العدمي)، والمنطقة (أو المناطق) خارج فترة الثقة confidence interval بمنطقة (أو مناطق) الرفض the region(s) of rejection (للفرض العدمي H_0) أو بالمنطقة (أو المناطق) الحرجة critical values.

ويمكن الربط بين أسلوب فترة الثقة confidence- interval. وأسلوب اختبار المعنوية test of significance لاختبارات الفروض من خلال مقارنة المعادلة (5.3.5) بالمعادلة (2.7.5).

ففي إجراء فترة الثقة، نحن نحاول إيجاد المدى أو الفترة التي يقع فيها المعلمة غير المعلومة β_2 باحتمال $(1-\alpha)$، بينما في أسلوب اختبار المعنوية نحن نفترض قيمة معينة للمعلمة غير المعلومة β_2 ، ونحاول أن نعرف هل القيمة المحسوبة للإحصاء $\hat{\beta}_2$ تقع داخل حدود الثقة أم لا.

ولتوضيح ذلك، سوف نعتبر مثال الدخل (الاستهلاك) حيث تم حساب $\hat{\beta}_2 = 0.5091$ ، $\text{se}\,(\hat{\beta}_2) = 0.0357$ ، بدرجات حرية $df = 8$، وعندما $\alpha = 0.05$ من جدول توزيع t نجد أن $t_{\alpha/2} = 2.306$.

فإذا اعتبرنا الفرض العدمي :

$$H_0 : \beta_2 = \beta_2^* = 0.3$$

والفرض البديل :

$$H_1 : \beta_2 \neq 0.3$$

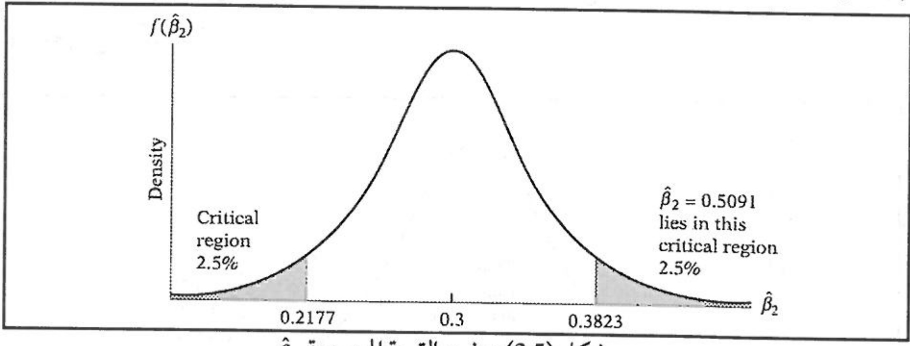
(انظر العلاقة (2.7.5)) فإن :

$$\Pr(0.2177 \leq \hat{\beta}_2 \leq 0.3823) = 0.95 \qquad (3.7.5)$$ [10]

وشكل (3.5) يوضح أن القيمة المحسوبة $\hat{\beta}_2$ حيث $\hat{\beta}_2 = 0.5091$ تقع في منطقة الرفض، لذلك نرفض الفرض العدمي القائل بأن $\beta_2 = 0.3$.

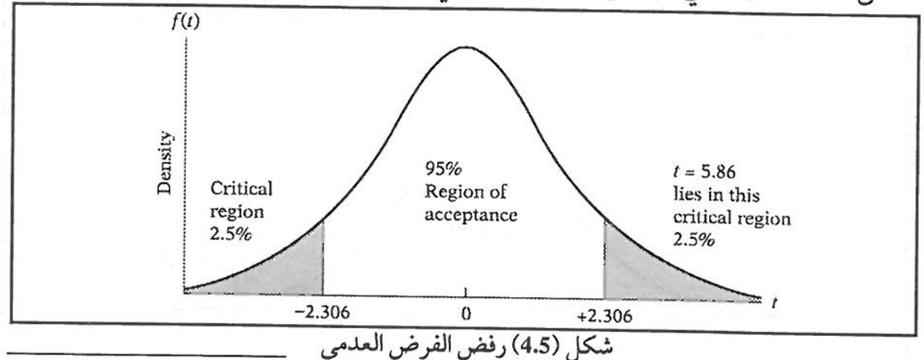
وعملياً، يمكننا حساب قيمة t حيث :

$$t = \frac{0.5091 - 0.3}{0.0357} = 5.86 \qquad (4.7.5)$$



شكل (3.5) يوضح القيمة المحسوبة $\hat{\beta}_2$

ونجد أن قيمة الإحصاء t تقع خارج حدود الثقة 2.306 ± كما هو موضح في شكل (4.5)، وبالتالي نرفض الفرض العدمي بدرجة ثقة 95% .



شكل (4.5) رفض الفرض العدمي

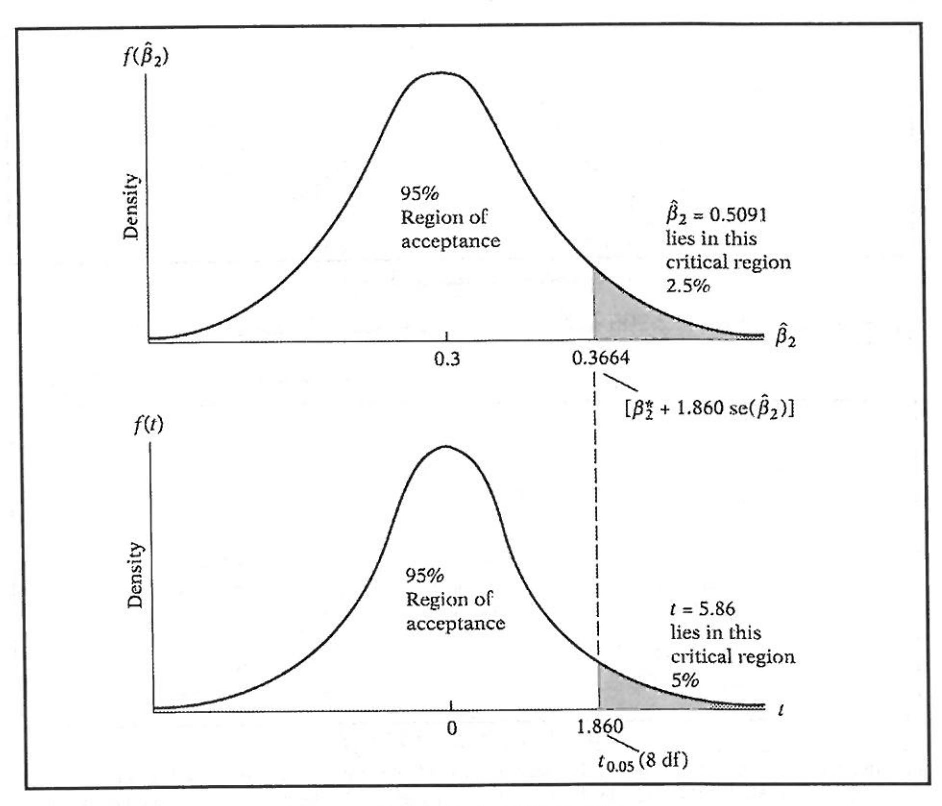
(10) In Sec. 5.2, point 4, it was stated that we cannot say that the probability is 95 percent that the fixed interval (0.4268. 0.5914) includes the true β_2. But we can make the probabilistic statement given in (5.7.3) because $\hat{\beta}_2$, being an estimator, is a random variable.

ومما هو جـدير بالذكـر (أنه عندما يكون تقـدير $\hat{\beta}_2$ أي $\hat{\beta}_2$ تساوي القيمة الافتراضية في الفرض العدمي ، فإن t في (4.7.5) سوف تساوي صفراً .

وكما سبق أن استخدمنا توزيع t في اختبار الذيلين two- sided or two- tail، حيث توجد منطقتان للرفض كما هو موضح بشكل (3.5) أو (4.5) حيث يأخذ الفرض البديل علاقة لايساوي .

ولكن لو فرضنا أن الخبرة السابقة تجعلنا نتوقع أن تكون قيمة β_2 أكبر من 0.3 .

في هذه الحالة، يكون الفرض العدمي في الشكل $H_0 : \beta_2 \leq 0.3$ والفرض العدمي $H_1 : \beta_2 > 0.3$ في هذه الحالة يسمى الاختبار اختبار الذيل الواحد جهة اليمين . وفي هذه الحالة تكون منطقة الرفض في الطرف الأيمن (اختبار الذيل الأيمن) هو موضح في شكل (5.5) .



شكل (5.5) اختبار الذيل الواحد

بالمثل في حالة إذا كان الفرض البديل في شكل $H_1: \beta_2 < 0.3$ – في هذه الحالة يكون الاختبار اختبار الذيل الواحد أيضاً ، وتكون منطقة الرفض في الطرف الأيسر اختبار الذيل الأيسر.

ومما سبق يمكن تلخيص اختبار t في جدول (1.5).

جدول (1.5) اختبار t للمعنوية : قواعد القرار

نوع الفرض	الفرض العدمي H_0	الفرض البديل $H_0 \neq \beta_2^*$	قاعدة القرار : رفض H_0 إلى I
الذيلين	$\beta_2 = \beta_2^*$	$\beta_2 \neq \beta_2^*$	$\lvert t \rvert > t_{a/2,\text{df}}$
الذيل الأيمن	$\beta_2 \leq \beta_2^*$	$\beta_2 > \beta_2^*$	$t > t_{a,\text{df}}$
الذيل الأيسر	$\beta_2 \geq \beta_2^*$	$\beta_2 < \beta_2^*$	$t > -t_{a,\text{d}}$

ملحوظة: β_2^* : هي القيمة العددية لـ β_2 المفترضة في الفرض العدمي

$\lvert t \rvert$: القيمة المطلقة لـ t مستوى المعنوية

$t_{a/2}$ أو t_a : تمثل حدود الثقة عند $\alpha/2$, α على التوالي

df : درجات الحرية $(n-2)$ بالنسبة لنموذج متغيرين، $(n-3)$ بالنسبة لنموذج الثلاثة متغيرات، هكذا. نفس الإجراء يمكن أن يتناول بالنسبة لاختبارات الفروض للمعلمة β_1.

اختبار المعنوية لـ σ^2:

Testing the significance of σ^2

اختبار X^2:

the X^2 test

ولإجراء اختبار عن المعلمة σ^2 (حيث σ^2 تمثل التباين للمتغير العشوائي u)، فإننا سوف نعتبر المتغير X^2 حيث :

$$\chi^2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \qquad (1.4.5)$$

ومما سبق ، يتضح أن X^2 متغير عشوائي له توزيع X^2 بدرجات حرية $(n-2)$.

فعلى سبيل المثال ، في مثال الدخل والاستهلاك نجد أن $\hat{\sigma}^2 = 42.1591$ بدرجات حرية df=8. فإذا فرضنا أن :

$$H_0 : \sigma^2 = 85 \ , \ H_1 : \sigma^2 \neq 85$$

والمعادلة (1.4.5) تمدنا بالتوزيع الاحتمالي للمتغير $\hat{\sigma}^2$ ، وباستخدام التوزيع الاحتمالي لـ X^2 وعند درجة ثقة $(1-\alpha)$ ودرجات حرية $(n-2)$ يمكننا إجراء الاختبارات لرفض أو عدم رفض الفرض العدمي $H_0 : \sigma^2 = \sigma_0^2$ والفرض البديل

$H_1: \sigma^2 \neq \sigma_0^2$ في حالة اختبار الذيلين، $H_1: \sigma^2 > \sigma_0^2$ في حالة اختبار الذيل الأيمن، أو $H_1: \sigma^2 < \sigma_0^2$ في حالة اختبار الذيل الأيسر.

وجدول (2.5) يلخص اختبار X^2 بالنسبة للمعلمة σ^2

جدول (2.5) تلخيص لاختبار X^2

الفرض العدمي H_0	الفرض البديل H_1	منطقة الرفض للفرض H_0 إلى I كان
$\sigma^2 = \sigma_0^2$	$\sigma^2 > \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha,df}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 < \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} < \chi^2_{(1-\alpha),df}$
$\sigma^2 = \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$	$\frac{df(\hat{\sigma}^2)}{\sigma_0^2} > \chi^2_{\alpha/2,df}$ or $< \chi^2_{(1-\alpha/2),df}$

ملحوظة: σ_0^2 ترمز إلى قيمة σ^2 تحت الفرض العدمي df ترمز لعدد درجات الحرية حيث $df = n-2$ إذا كان النموذج في متغيرين، كذلك $df = n-3$ إذا كان النموذج في 3 متغيرات.

8.5 اختبارات الفروض.. بعض الجوانب العملية:
HYPOTHESIS TESTING: SOME PRACTICAL ASPECTS

معنى قبول أو رفض الفرض: The Meaning of "Accepting" or "Rejecting" a Hypothesis

إذا اعتبرنا اختبار المعنوية أي اختبار t، وتم قبول "accept" الفرض العدمي، ولكن عادة نقول إنه على أساس بيانات العينة، فإنه لايوجد سبب لرفض الفرض العدمي بدلاً من القول إن الفرض العدمي صحيح وذلك منعاً للشك.

ولتوضيح ذلك، دعنا نعود إلى مثال الدخل والاستهلاك، فإذا كان الفرض العدمي $H_0: \beta_2 = 0.50$، وأن $\hat{\beta}_2 = 0.5091$. $se(\hat{\beta}_2) = 0.0357$ فإنه على أساس اختبار t نجد أن:

$$t = \frac{(0.5091 - 0.50)}{0.0357} = 0.25$$

وبما أن $t = 0.25$ فهذا يعني أن الفرض العدمي غير معنوي insignificant، عند درجة ثقة 95%. لذلك فإنه يمكن قبول الفرض H_0، ولكن إذا فرضنا أن $H_0 = 0.48$ ونحسب t نجد أن:

$$t = \frac{(0.5091 - 0.48)}{0.0357} = 0.82$$

t و في هذه الحالة أيضاً غير معنوية insignificant ، وفي هذه الحالة أيضاً تقبل "accept" الفرض العدمي ، وهنا يطرح السؤال التالي نفسه أي الفرضين $H_0 = 0.5$ أو $H_0 = 0.48$ صحيح . ولايمكن الإجابة حالياً على هذا السؤال . فقبول الفرض العدمي دائماً محاط بفروض أخرى يمكن قبولها ، لذلك يكون من الأفضل القول "إننا يمكن قبول الفرض العدمي أو لايوجد سبب لرفض الفرض العدمي"[11].

الفرض العدمي "صفر" وقاعدة ذامب :

The "Zero" Null Hypothesis and the "2-t" Rule of Thumb

إذا كان الفرض العدمي $H_0: \beta_2 = 0$ فهذا الفرض العدمي " صفر " يعني اختبار هل توجد علاقة خطية بين المتغير التابع Y والمتغير المستقل X. ويمكن إجراء الاختبار بأسلوب فترة الثقة the confidence interval أو أسلوب اختبار t the t- test approach، ويسمى الاختبار في هذه الحالة بقاعدة ذامب rule of thumb.

قاعدة ذامب **Thumb of rule** : إذا كان عدد درجات الحرية df 20 فأكثر، ومستوى المعنوية $\alpha = 0.05$ فإن الفرض العدمي $\beta_2 = 0$ يتم رفضه إذا كانت قيمة t المحسوبة من العلاقة (2.3.5) : $$t = \hat{\beta}_2 \ / \ se(\hat{\beta}_2)$$ تزيد عن القيمة الموجبة absolute value لـ 2

وباختصار ، فإنه يمكن رفض الفرض العدمي $H_0: \beta_2 = 0$ إذا كان :

$$t = \hat{\beta}_2/\text{se}(\hat{\beta}_2) > t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 > 0$$

أو :

$$t = \hat{\beta}_2/\text{se}(\hat{\beta}_2) < -t_{\alpha/2} \qquad \text{when } \hat{\beta}_2 < 0$$

أو بعبارة أخرى :

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)}\right| > t_{\alpha/2} \qquad \textbf{(1.8.5)}$$

وفي حالة اختبار الذيل الواحد عندما $H_0: \beta_2 = 0$، $H_1: \beta_2 > 0$ أو $H_1: \beta_2 < 0$ فإننا نرفض الفرض العدمي إذا كان :

$$|t| = \left|\frac{\hat{\beta}_2}{\text{se}(\hat{\beta}_2)}\right| > t_{\alpha} \qquad \textbf{(2.8.5)}$$

(11) Jan Kmenta, Elements of Econometrics, Macmillan, New York, 1971, p. 114.

صياغة الفرض العدمي والفرض البديل[12]:

Forming the null and alternative hypothesis

وهنا يطرح سؤال : كيف تتم صياغة الفرض العدمي والفرض البديل؟ وهنا لاتوجد قواعد محددة يجب اتباعها ، ولكن عادة تتم الصياغة للفرض العدمي والبديل بناء على طبيعة الظاهرة محل الدراسة. فعلى سبيل المثال، إذا اعتبرنا خط سعر رأس المال (CML) لنظرية حافظة الأوراق المالية portfolio theory، التي تفترض :

$$E_i = \beta_1 + \beta_2 \sigma_i$$

حيث : E تساوي قيمة العائد المتوقع على الحافظة، σ تساوي الانحراف المعياري للعائد، ومقياس للمخاطرة أيضاً. وبما أن العائد والمخاطرة يتوقع أن العلاقة بينهما علاقة طردية (أي علاقة موجبة)، بمعنى كلما زاد العائد زادت المخاطرة والعكس صحيح.

ففي هذه الحالة ،يكون طبيعيًا أن يكون الفرض العدمي $H_0: \beta_2 = 0$ والبديل $H_1: \beta_2 > 0$

ففي هذه الحالة يكون من غير المنطقي اعتبار $\beta_2 < 0$.

وهذا يعني أن صياغة الفرض العدمي والبديل ترجع إلى طبيعة الظاهرة أو المشكلة محل الدراسة. ومثال آخر هو الطلب على النقود وكما ذكرنا سابقاً، من المحددات المهمة للطلب على النقود هو الدخل قبل الدراسة لدوال الطلب النقدية، حيث توضح الدراسة أن المرونة الدخلية income elasticity للطلب على النقود (أي نسبة التغير في الطلب للنقود لتغير الدخل نسبة 1%) وعادة تكون محصورة بين 0.7 ، 1.3. لذلك في دراسة الطلب على النقود، فإنه يمكن افتراض أن المعامل β_2 الذي يمثل المرونة الدخلية يساوي 1، والفرض البديل $\beta_2 \neq 1$ أي استخدام اختبار الذيلين.

ومما سبق ، نخلص إلى أن التوقعات النظرية أو الحوادث التكرارية السابقة يمكن أن تمكننا من صياغة الفرض العدمي والبديل.

اختيار مستوى المعنوية α : Choosing α, the Level of Significance

ومما سبق ، يتضح أن رفض أو عدم رفض الفرض العدمي ، يعتمد على القيم الحرجة لمستوى المعنوية α أو احتمال الخطأ من النوع الأول type 1 error وهو يعني

---

(12) For an interesting discussion about formulating hypotheses, see J. Bradford De Long and Kevin Lang, "Are All Economic Hypotheses False?" Journal of Political Economy, vol, 100, no. 6, 1992, pp. 1257-1272.

احتـمـال رفض فـرض صـحـيح. وفي ملحق A نتناول بـالتـفـصـيل الخطأ من النوع الأول، وعلاقته بـالخطأ من النوع الثاني type II error (احـتـمـال قبـول فـرض خطأ) كذلك لماذا تركز الإحصاءات التقليدية على الخطأ من النوع الأول. كذلك لماذا قيمة α عادة تساوي 1، 5 أو على الأكثر 10 في المائة؟ وفي ملحق A يتضح أنه عند حجم معين إذا تم تخـفـيض احـتـمـال الخطأ من النوع الأول (α) فإن هذا سوف يؤدي إلى زيادة احتمال الخطأ من النوع الثاني II والعكس صحيح.

وهذا يعني أنه عند حجم عينة معين إذا تم تخفيض احتـمـال رفض عـدمي صـحـيح، فإن هذا سوف يؤدي إلى زيادة احتـمـال قبـول فـرض عـدمي خطأ. وهذا يعني أنه عند حجم معين للعينة، فإنه توجد علاقة تبادلية بين احتمال الخطأ من النوع الأول واحـتـمـال الخطأ من النوع الثاني. وهنا يتضح أن الأسلوب الوحيد لتحديد احـتـمـال الخطأ من النوع الأول وبالتـالي الثاني أيضاً يعتمد على التكاليف النسبية relative costs لكل نوع من الأخطاء. فإذا كان خطأ لرفض فـرض عـدمي صـحـيح (النوع الأول من الخطأ) مرتبط نسبياً بخطأ عدم رفض فرض عـدمي خطأ (أي الخطأ من النوع الثاني). وعادة في الواقع تكون تكلفة الخطأ من النوع الأول أقل من تكـلفة الخطأ من النوع الثاني[13].

لذا بصفة عامة، يختار المتخصصون في الاقتصاد القياسي قيمًا لاحتمال الخطأ من النوع الأول (α) قيم 1 أو 5 أو 10 على الأكثر بحيث يكون احتمال الخطأ من النوع الثاني أقل مايمكن. ويسمى الاحتمال المكمل لاحتمال الخطأ من النوع الثاني بقـوة الاختبار power of the test حيث إن قوة الاختبار تعني احتمال قبول فرض صحيح. وبالتالي فعادة أي إجراء يهدف إلى تعظيم قوة الاختبار (توجد تفصيلات عن قوة الاختبار بملحق A). ويمكننا تجنب مشكلة تحديد القيمة المثلى لـ α وذلك بتحديد قيمة *p* value *p* بالنسبة لإحصاء الاختبار the test statistic وهو ماسوف نتناوله فيما بعد.

المستوى التام للمعنوية.. قيمة *p*:

The Exact Level of Significance: The *p* Value

تعرف قيمة *p* أو *p*-value بأنها أقل احتمال لرفض الفرض العدمي (أي أقل احتمال لرفض فرض عدمي صحيح) the lowest significance level at which a null hypothesis can be rejected

(13) Jan Kmenta, Elements of Econometrics, Macmillan, New York, 1971, pp. 126–127.

ولتوضيح ذلك ، إذا اعتبرنا مثال الدخل والاستهلاك ، فإذا كان الفرض العدمي أن قيمة MPC تساوي 0.3. فإذا كانت قيمة إحصاء الاختبار t حيث $t = 5.86$ في المعادلة (4.7.5) .

وبالتالي ، باستخدام الجدول الإحصائي لتوزيع t عند درجات حرية 8 (df= 8) بملحق D نجد أن احتمال أن تكون t أكبر من $0.000189 \approx 0.0002$[14] أي أن $p \approx 0.0002$. وقيمة $p \approx 0.0002$ يعني أن أقل احتمال لرفض فرض صحيح تساوي تقريباً 0.0002 وهذه القيمة تناظر أكبر احتمال لقبول فرض خطأ (أي الاحتمال المكمل لاحتمال الخطأ من النوع الثاني) .

المعنوية الإحصائية مقابل المعنوية العملية :

Statistical Significance versus Practical Significance

بالرجوع إلى مثال الدخل والاستهلاك ، فإذا كان الفرض العدمي $H_0: \beta_2 = 0.61$ ، فإذا كان تقدير قيمة β_2 من بيانات العينة تساوي 0.5091 أي أن $\hat{\beta}_2 = 0.5091$ ، وبالتالي فإن فترة الثقة لـ β_2 بناء على بيانات العينة تكون (0.4268 ، 0.5914) بدرجة ثقة 95% ، وبالتالي فإن قيمة $\beta_2 = 0.61$ في الفرض العدمي تقع خارج فترة الثقة ، وبالتالي فإنه معنوي إحصائياً ، وبالتالي $\beta_2 = 0.61$ تكون معنوية إحصائياً statically significant بدرجة ثقة 95% .

ولكن ماهي المعنوية العملية practical significance؟ أو ماهو الفرق بين أن يكون MPC= 0.61 بدلاً من MPC= 0.5091؟ أو بعبارة أخرى الفروق (0.1009) بين قيمتي MPC تعتبر مهمة عملياً أم لا؟

وإجابة هذا السؤال تعتمد على كيفية استخدام تقدير β_2. فعلى سبيل المثال، معروف في الاقتصاد الكلي macroeconomics أن معامل الدخل the income multiplier يساوي من 1/(1–MPC) ، وبالتالي فإذا كان MPC = 0.5091 فإن معامل الدخل يساوي 2.04 ويساوي 2.56 إذا كان MPC = 0.61 . فإذا رفعت الحكومة الإنفاق بدولار واحد لخروج الاقتصاد من الانكماش ، فإنه في النهاية سوف يؤدي إلى زيادة الدخل بـ 2.04 دولار ، وذلك عند MPC = 0.5091 . أما إذا كانت قيمة MPC بحيث MPC = 0.61 فإن ذلك سوف يؤدي إلى زيادة الدخل إلى 2.56 عند زيادة الإنفاق

(14) One cna obtain the p value using electronic statistical tables to several decimal places. Unfortunately, the conventional statistical tables, for lack of space, cannot be that refined Most statistical packages now routinely print out the p values.

بدولار واحد. وبالتالي الزيادة في الدخل من 2.04 إلى 2.56 دولار سوف يؤدي إلى سلسلة من الانتعاش الاقتصادي.

ومما سبق، يتضح أنه لايوجد تعارض بين المعنوية الإحصائية statistical significance والمعنوية العملية أو الاقتصادية practical or economic significance. وفي هذه النقطة يمكن الرجوع إلى عالم الاقتصاد goldberger[15].

الاختيار بين أسلوبي فترة الثقة واختبار المعنوية لاختبار الفرض:

The choice between confidence interval and test of significance approaches to hypothesis testing

في معظم التحليلات الاقتصادية التطبيقية يتم وضع قيمة صغيرة لـ β_2 في الفرض العدمي، مما يؤدي عند إجراء الاختبار إلى رفض الفرض العدمي. فمثلاً في مثال الدخل والاستهلاك يفترض أن الفرض العدمي H_0: MPC $\beta_2 = 0$ فافتراض أن $\beta_2 = 0$ رغم أنها قيمة مرفوضة، إلا أنه يتم استخدامها لتفسير بعض النتائج التكرارية.

وعادة يفضل كثير من الكتاب استخدام أسلوب فترة الثقة the confidence interval approach بدلاً من the test of significance approach. ويترك للقارئ الأسلوب الذي يتم استخدامه[17].

9.5 تحليل الانحدار وتحليل التباين:
REGRESSION ANALYSIS AND ANALYSIS OF VARIANCE

في هذا الفصل، سوف نتناول بالدراسة تحليل الانحدار من جهة تحليل الانحدار analysis of variance، وذلك لتقدم للقارئ طريقة تكميلية للبحث في مشكلة الاستدلال الإحصائي.

في الفصل الثالث، (5.3)، تم تحديد العلاقة التالية:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum \hat{u}_i^2 = \hat{\beta}_2^2 \sum x_i^2 + \sum \hat{u}_i^2 \qquad (3.5.2)$$

(15) Arthur S. Goldberger, A Course in Econometrics, Harvard University Press, Cambridge, Massachusetts, 1991, p. 240. Note bi is the OLS estimator of bj and $\hat{\sigma}_{bj}$ is its standard error. For a corroborating view, see d. N. McCloskey, "The Loss Function Has Been Mislaid: The Rhetoric of Significance Tests", American Economic Review, vol. 75, 1985, pp. 201-205. See also D. N. McCloskey and S. T. Ziliak, "The Standard Error of Regression," Journal of Economic Literature, vol. 37, 1996, pp. 97–114.

(16) See their article cited in footnote 12, p. 1271.

(17) For a somewhat different perspective, see Carter Hill, William Griffiths, and George Judge, Undergraduate Econometrics, Wiley & Sons, New York, 2001, p. 108.

أو بعبارة أخرى :

$$TSS = ESS + RSS$$

وهذه العلاقة توضح تقسيم المجموع الكلي للمربعات the total sum of squares (TSS) إلى مجموعتين من المكونات هما: المجموع المفسر للمربعات explained sum of squares (ESS) ومجموع البواقي للمربعات residual sum of squares (RSS). ودراسة المكون (TSS) تسمى بتحليل التباين the analysis of variance وللاختصار يرمز له بالرمز (ANOVA) للانحدار.

ودائماً مجموع المربعات يرتبط بعدد درجات الحرية (df)، حيث تمثل عدد درجات الحرية عدد المشاهدات المستقلة. ونجد أن درجات الحرية المناظرة لـ TSS تساوي ($n-1$) كذلك عدد درجات الحرية المناظرة لـ RSS تساوي ($n-2$). (ويعتبر ذلك صحيحًا بالنسبة لنموذج الانحدار في متغيرين)، كذلك المجموع ESS له درجة حرية تساوي واحداً. يسمى الجدول الذي يحتوي على مكونات المجاميع ودرجات الحرية المناظرة لها بجدول AOV أو مايسمى بجدول ANOVA كما هو موضح بجدول (3.5). ووفقاً لمدخلات جدول (3.5) يمكن تعريف المتغير التالي :

$$\begin{aligned} F &= \frac{\text{MSS of ESS}}{\text{MSS of RSS}} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\sum \hat{u}_i^2/(n-2)} \qquad \textbf{(1.9.5)} \\ &= \frac{\hat{\beta}_2^2 \sum x_i^2}{\hat{\sigma}^2} \end{aligned}$$

وإذا فرضنا أن المتغير العشوائي U_i يتبع التوزيع المعتاد (الطبيعي)، وإذا كان الفرض العدمي (H_0) بحيث $\beta_2 = 0$، فإنه يتضح أن المتغير F في (1.9.5) يتبع توزيع فيشر F بدرجات حرية في البسط تساوي واحداً ودرجات حرية المقام تساوي ($n-2$) (انظر ملحق 3A5 في الفصل (5) للإثبات، بالإضافة إلى الخصائص العامة لتوزيع F، كما هو موضح بملحق A. وهنا يطرح السؤال التالي، فيما تستخدم النسبة F؟ ويمكن توضيح ذلك على النحو التالي[18] :

$$E\left(\hat{\beta}_2^2 \sum x_i^2\right) = \sigma^2 + \beta_2^2 \sum x_i^2 \qquad \textbf{(2.9.5)}$$

(18) For proof, see K. A. Brownlee, Statistical Theory and Methodology in Science and Engineering, John Wiley & Sons, New York, 1960, pp. 278–280.

$$E\frac{\sum \hat{u}_i^2}{n-2} = E(\hat{\sigma}^2) = \sigma^2 \qquad (3.9.5)$$

(ويلاحظ أن القيم الفعلية للمعلمات β_2، σ^2 تظهر في الطرف الأيمن للمعادلتين السابقتين).

وبالتالي إذا كانت β_2 تساوي صفراً ، فإن الطرف الأيمن في المعادلتين (2.9.5), (3.9.5) يكون متساويًا. وفي هذه الحالة ، لايكون للمتغير المفسر X تأثير خطي على Y ، ويرجع الاختلاف في قيم Y إلى المتغير العشوائي u_i. ومن جهة أخرى ، إذا كانت قيمة β_2 تختلف عن الصفر ،فإن الطرف الأيمن في (2.9.5), (3.9.5) يختلف عن σ^2 بمقدار الاختلاف في Y الراجع إلى التغير في X. كذلك النسبة F في (1.9.5) تمدنا برفض أو عدم رفض الفرض العدمي $H_0: \beta_2 = 0$. ولتوضيح ذلك يمكن اعتبار مثال الدخل والاستهلاك، فإن جدول ANOVA موضح بجدول A5، ومن الجدول نجد أن $F = 202.87$، وقيمة p المناظرة لقيمة F عند درجات حرية واحد (في البسط)، 8 (في المقام) من الدوال الإحصائية لتوزيع F في ملحق D نجد أن p تساوي 0.0000001. وإذا تم اختيار أسلوب مستوى المعنوية فإن $\alpha = 0.01$، ونجد أن قيمة F المحسوبة أي $F = 202.87$ تكون معنوية عند مستوى المعنوية $\alpha = 0.01$.

لذلك إذا تم رفض الفرض العدمي $H_0: \beta_2 = 0$ فإن احتمال الخطأ من النوع I يكون صغيراً جداً.

جدول (4.5)

Source of variation	SS	df	MSS	
Due to regression (ESS)	8552.73	1	8552.73	$F = \frac{8552.73}{42.159}$
Due to residuals (RSS)	337.27	8	42.159	$= 202.87$
TSS	8890.00	9		

ومن الجانب العملي، بما أن العينة في مثال الدخل والاستهلاك مسحوبة من مجتمع ، حيث $\beta_2 \neq 0$ لذلك نحن نستنتج بدرجة ثقة كبيرة أن الدخل X يؤثر على الإنفاق الاستهلاكي Y.

وبالرجوع لنظرية (5–7) في ملحق (1.A5) التي تقر أن مربع قيمة t بدرجات حرية k تساوي القيمة F بدرجات حرية واحد (للبسط)، ودرجات حرية تساوي k (للمقام).

وبالنسبة لمثال الدخل والاستهلاك ، إذا فرضنا $H_0: \beta_2 = 0$ فمن العلاقة (2.3.5) يمكن بسهولة إثبات أن القيمة التقديرية t تساوي 14.26 بدرجات حرية تساوي 8 .

وتحت نفس الفرض العدمي ، فإن القيمة $F = 202.87$ بدرجات حرية تساوي واحد، 8 . لهذا فإن القيمة $(14.24)^2$ تساوي القيمة F .

هكذا، فإن كلاً من اختبار t، واختبار F يعتبران اختبارين بديلين يستخدمان لاختبار الفرض $\beta_2 = 0$. وبالنسبة لنموذج الانحدار في متغيرين عادة يستخدم اختبار t ولانحتاج في الغالب إلى استخدام اختبار F. أما بالنسبة لنماذج الانحدار متعددة المتغيرات فإنه يتم استخدام اختبار F في العديد من التطبيقات المهمة .

10.5 تطبيق تحليل الانحدار.. مشكلة التنبؤ :

APPLICATION OF REGRESSION ANALYSIS: THE PROBLEM OF PREDICTION

على أساس عينة البيانات في جدول (2.3)، فإنه تم الحصول على نموذج الانحدار التالي :

$$\hat{Y}_i = 24.4545 + 0.5091X_i \tag{2.6.3}$$

حيث $\hat{Y}_1$ هي تقدير القيمة الفعلية $E(Y_i)$ عند قيمة معينة X_i.

والآن ، سوف نوضح كيفية استخدام الانحدار التاريخي historical regression؟ أحد استخدامات الانحدار التاريخي هو استخدامه في التنبؤ prediction أو forcasting . ففي مثال الدخل والاستهلاك، فإن الإنفاق الاستهلاكي في فترة زمنية مستقبلية يمكن الحصول عليه عند مستوى دخل معين X.

ويوجد نوعان من التنبؤ هما :

1 - التنبؤ بالقيمة المتوقعة الشرطية لـ Y عند مستوى معين لـ X وليكن X_0، وهي نقطة تقع على خط انحدار المجتمع نفسه (انظر شكل 2.2) .

2 - التنبؤ بقيمة Y المشاهدة وفقاً لمستوى الدخل X_0. ويسمى النوع الأول من التنبؤ بالتنبؤ المتوسط mean prediction ، ويسمى التنبؤ الثاني بالتنبؤ المشاهد individual prediction .

التنبؤ المتوسط : Mean prediction[19]

إذا فرضنا أن $X_0 = 100$ ونرغب في التنبؤ بالقيمة المتوسطة (أو المتوقعة) لـ Y بشرط $X_0 = 100$ أي نرغب في التنبؤ بـ $E(Y|X_0 = 100)$.

ويمكن توضيح الانحدار التاريخي في (2.6.3) يمثل تقدير نقطة لمتوسط التنبؤ على النحو التالي :

$$\begin{aligned}\hat{Y}_0 &= \hat{\beta}_1 + \hat{\beta}_2 X_0 \\ &= 24.4545 + 0.5091(100) \\ &= 75.3645\end{aligned} \quad (1.10.5)$$

حيث ترمز $\hat{Y}_0$ بتقدير $E(Y|X_0)$. ويمكن إثبات أن تنبؤ النقطة السابقة أفضل من التقدير غير المتحيز الخطي الأمثل best linear unbiased estimator (BLUE) وبما أن Y_0 تقدير، بالتالي فإنه يمكن أن يختلف عن القيمة الفعلية Y_0. والفرق بين القيمة الفعلية Y_0 والقيمة التقديرية $\hat{Y}_0$، وهذا الاختلاف يسمى بخطأ التنبؤ prediction or forecast error. ولتحديد الخطأ، نحن نحتاج إلى إيجاد توزيع المعاينة للتقدير $\hat{Y}_0$. ومن ملحق 4A5 فصل (5). نجد أن $\hat{Y}_0$ متغير عشوائي يتبع التوزيع المعتاد (الطبيعي) بتوقع $(\beta_1+\beta_2X_0)$ والتباين التالي :

$$\operatorname{var}(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \quad (2.10.5)$$

وعند إحلال المعلمة σ^2 بتقديرها غير المتحيز $\hat{\sigma}^2$ فإن المتغير التالي :

$$t = \frac{\hat{Y}_0 - (\beta_1 + \beta_2 X_0)}{\operatorname{se}(\hat{Y}_0)} \quad (3.10.5)$$

متغير يتبع توزيع استيودنت (t) بدرجات حرية $(n-2)$. وبالتالي، فإنه يمكن استخدام توزيع t لاشتقاق فترات الثقة بالنسبة لقيمة $E(Y_0|X_0)$ واختبارات الفروض حولها على النحو :

$$\Pr[\hat{\beta}_1 + \hat{\beta}_2 X_0 - t_{\alpha/2}\operatorname{se}(\hat{Y}_0) \le \beta_1 + \beta_2 X_0 \le \hat{\beta}_1 + \hat{\beta}_2 X_0 + t_{\alpha/2}\operatorname{se}(\hat{Y}_0)] = 1 - \alpha \quad (4.10.5)$$

حيث يتم الحصول على $\operatorname{se}(\hat{Y}_0)$ من المعادلة (2.10.5). ومن بيانات في جدول (3-3) نجد أن :

(19) For the proofs of the various statements made, see App. 5A, Sec. 5A. 4.

$$\text{var}(\hat{Y}_0) = 42.159\left[\frac{1}{10} + \frac{(100-170)^2}{33{,}000}\right]$$
$$= 10.4759$$

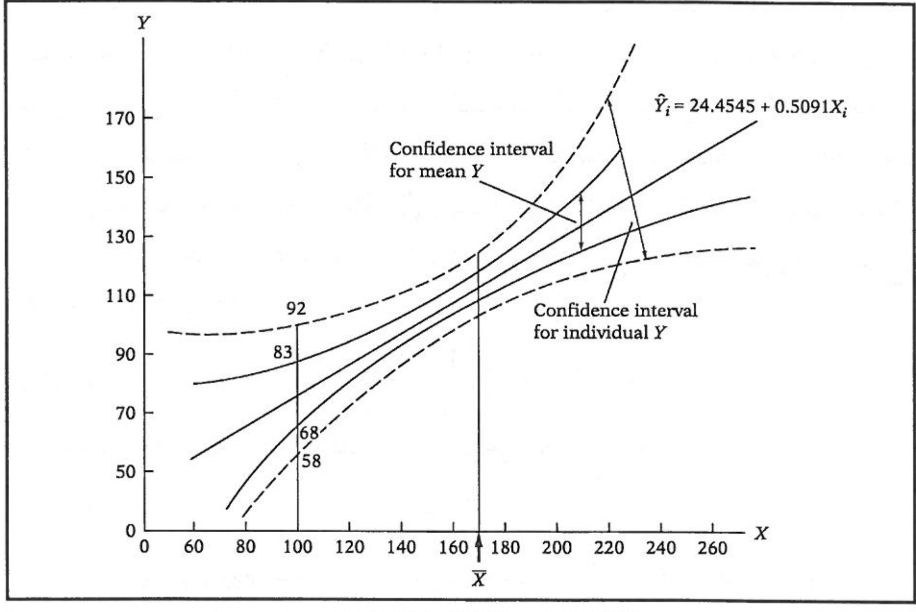
و :
$$\text{se}(\hat{Y}_0) = 3.2366$$

وعند درجة ثقة 95% أي $1-\alpha = 0.95$ نجد أن فترة القيمة الفعلية لـ $E(Y|X_0) = \beta_1 + \beta_2 X_0$ على النحو

$$75.3645 - 2.306(3.2366) \le E(Y_0|\, X = 100) \le 75.3645 + 2.306(3.2366)$$

أي أن

$$67.9010 \le E(Y \mid X = 100) \le 82.8381 \qquad \textbf{(5.10.5)}$$



شكل (6.5) فترة الثقة لدالة انحدار المجتمع

ويتضح من شكل (6. 5) أنه عند $X_0 = 100$ في 100 عينة تكرارية يكون في 95 منها قيمة $E(Y|X = 100)$ الفعلية تقع داخل الفترة في العلاقة (5.10.5). وأن أفضل تقدير للقيمة المتوقعة هو تقدير النقطة ، وهو يقع على خط الانحدار عند $X_0 = 100$ فإن $\hat{Y}_0 = 75.3645$. وفي حالة الحصول على 95% من فترات الثقة كما في (5.10.5) لكل قيمة من قيم X في جدول (2.3) نحن نحصل على ماهو معروف بفترة الثقة confidence interval لدالة انحدار المجتمع the population regression function كما هو موضح في شكل (6.5).

التنبؤ بقيمة المشاهدة: Individual Prediction

وإذا كان اهتمامنا ينحصر في التنبؤ بقيمة Y المشاهدة ولتكن Y_0 عند قيمة محددة لـ X ولتكن X_0، فكما يتضح من ملحق 3.A5 الفصل (5)، أن أفضل تقدير خطي غير متحيز Y_0 كما هو موضح في العلاقة (1.10.5)، وتباينه على النحو

$$\operatorname{var}(Y_0 - \hat{Y}_0) = E[Y_0 - \hat{Y}_0]^2 = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \qquad (6.10.5)$$

وكما سوف يتضح فيما بعد أن Y_0 متغير عشوائي يتبع التوزيع المعتاد (الطبيعي) بتوقع وتباين كما في (6.10.5) ، (1.10.5) على الترتيب. وبإحلال $\hat{\sigma}^2$ فإن المتغير :

$$t = \frac{Y_0 - \hat{Y}_0}{\operatorname{se}(Y_0 - \hat{Y}_0)}$$

متغير يتبع توزيع t. لذلك يمكن استخدام توزيع t لعمل استدلال حول القيمة الفعلية لـ Y_0.

وبالرجوع إلى مثال الدخل والاستهلاك، نحن نرى أن تنبؤ النقطة لـ Y_0 يساوي 75.3645 وهي نفس القيمة لـ $\hat{Y}_0$ بتباين يساوي 52.6349 (يمكن للقارئ إثبات هذا الاستنتاج). لذلك فإن فترة الثقة 95% لـ Y_0 وفقاً لـ $X_0 = 100$ تصبح على النحو :

$$(58.6345 \le Y_0 \mid X_0 = 100 \le 92.0945) \qquad (7.10.5)$$

وبمقارنة فترة الثقة أعلاه بفترة الثقة في (5.10.5) نحن نرى أن فترة الثقة للتنبؤ المشاهد Y_0 تكون أكبر من القيمة المتوسطة لـ Y_0 (لماذا)؟.

وبحساب فترات الثقة كما في العلاقة (5.10.7) بشرط قيم X المعطاة في جدول (2.3)، نحن نحصل على فترة ثقة 95% للقيم المشاهدة لـ Y وفقاً للقيم المعطاة لـ X، وشكل (6.5) يوضح فترة الثقة هذه مع فترة الثقة لـ $\hat{Y}_0$ عند قيم X المختلفة ويتضح من شكل (6.5) أن الفرق بين فترات الثقة يكون أقرب مايمكن عن بعضهم عندما $X_0 = X$ (لماذا)؟.

كذلك يزداد الفرق كلما تحركت X_0 بعيداً عن $\bar{X}$ (لماذا)؟ ، ومن ذلك يتضح أن الفترة التنبؤية من انحدار العينة التاريخي the historical sample regression line تفشل عندما X_0 تبتعد عن $X_0 = \bar{X}$.

لذلك يكون من الأهمية التفرقة بين خط الانحدار التاريخي للتنبؤ $E(Y|X_0)$، Y_0 عند قيمة معينة لـ X_0 عندما تبتعد X_0 عن $\bar{X}$.

11.5 تقرير نتائج تحليل الانحدار:
REPORTING THE RESULTS OF REGRESSION ANALYSIS

توجد طرق متعددة لتقرير نتائج تحيل الانحدار، ولكن في هذا المرجع ، سوف نستخدم الصياغة التالية السابق استخدامها في مثال الدخل والاستهلاك في الفصل (3) على النحو الموضح :

$$\begin{array}{llll} \hat{Y}_i = 24.4545 & + \quad 0.5091X_i & \\ \text{se} = (6.4138) & (0.0357) & r^2 = 0.9621 \\ t = (3.8128) & (14.2605) & \text{df} = 8 \\ p = (0.002571) & (0.000000289) & F_{1,8} = 202.87 \end{array}$$

في المعادلات (1.11.5) الأرقام بين القوسين في سطر se تمثل الأخطاء المعيارية لمعاملات الانحدار، وتمثل الأرقام في سطر t تقدير قيم t المحسوبة من المعادلة (2.3.5). وبتحديد قيمة p المناظرة لـ t عند مستوى معنوية معين α. فتحت الفرض العدمي H_0: $\beta_1 = 0$ فإن $p = 0.0026$ التي تشير إلى احتمال الحصول على t تساوي أو أكبر من 3.8128. لذلك، إذا رفضنا الفرض العدمي H_0، فإن احتمال وقوعنا في الخطأ من النوع الأول يساوي 0.0026%. وعملياً، يمكن القول إن β_1 تختلف عن الصفر. بنفس الطريقة، يمكن تفسير قيمة P بالنسبة لمعامل الانحدار β_2 تحت الفرض العدمي H_0: $\beta_2 = 0$.

وكما أوضحنا سابقاً العلاقة بين الإحصائيين F, t بمعنى أن $F_{1,k} = t_k^2$. وتحت الفرض العدمي أن $\beta_2 = 0$، العلاقة (1.11.5) توضح أن F حيث $F = 202.87$ (حيث درجة حرية واحدة للبسط، وعدد 8 من درجات الحرية للمقام) كذلك $t = 14.24$ بدرجات حرية $k = 8$.

12.5 تقييم نتائج تحليل الانحدار :
EVALUATING THE RESULTS OF REGRESSION ANALYSIS

في المقدمة شكل (4) يوضح الخطوات التصاعدية للاقتصاد القياسي للنمذجة. والآن سوف نقدم نتائج تحليل الانحدار لمثال الدخل والاستهلاك في (1.11.5)، وهنا يطرح سؤال عن مدى ملاءمة توفيق النموذج the fitted model أو بعبارة أخرى ماهي جودة توفيق النموذج؟ وبالتالي نحتاج إلى معيار criteria ما للإجابة على هذا السؤال.

أولاً: هل إشارات معاملات الانحدار المقدرة تتوافق مع النظرية أو التوقعات السابقة أم لا؟ فمعلوم أن β_2 تمثل المعدل الهامشي للاستهلاك (MPC) في دالة الاستهلاك، وتكون موجبة، وهذا يتفق مع المثال.

ثانياً: تعتبر النظرية أن العلاقة بين الدخل والاستهلاك ليست موجبة فقط، بل معنوية إحصائياً statistically significant، وبالتالي هل هذا محقق في المثال؟ وكما سبق توضيح أن MPC ليست فقط موجبة، بل أيضاً معنوية إحصائياً في فصل (11.5) وتختلف عن الصفر، وأن قيمة p صغيرة جداً.

ثالثاً: لم يفسر نموذج الانحدار من الاختلافات في الإنفاق الاستهلاكي الراجعة إلى الاختلافات في الدخل؟ وهنا يمكن استخدام r^2 للإجابة على هذا السؤال. ففي المثال وجدنا أن $r^2 = 0.96$ وهي تمثل نسبة كبيرة تقترب من الواحد، وهذا يعني أن معظم التغير في الإنفاق الاستهلاكي يرجع إلى التغير في الدخل. هكذا يتضح أن اختبار النموذج لشرح الإنفاق الاستهلاكي اختيار جيد تماماً quite good والآن يجب توضيح هل النموذج الذي تم بناؤه يحقق فروضًا للنموذج CNLRM. والآن سوف نوضح كيفية اختبار الفرض الطبيعي (العادي) أي اختبار هل المتغير العشوائي u_i يتبع التوزيع الطبيعي (أو المعتاد) أم لا؟ فاستخدام اختبار t أو F يتطلب ذلك تحقيق الفرض الطبيعي (العادي)، ففي حالة عدم تحقق هذا الفرض، أي عدم اتباع المتغير العشوائي u_i للتوزيع الطبيعي، فإنه لايمكن استخدام اختبار t أو F بالنسبة للعينات صغيرة الحجم وذات الحجم المحدد small or finite samples.

اختبارات فرض التوزيع الطبيعي: Normality Tests

توجد اختبارات متعددة لاختبار الفرض الطبيعي، أي اختبار هل المتغير العشوائي u_i يتبع التوزيع الطبيعي (المعتاد) أم لا. وفيما يلي سوف نقدم ثلاثة اختبارات فقط للفرض الطبيعي وهي: (1) اختبار المدرج التكراري للبواقي histogram of residuals، (2) رسم دالة الاحتمال للمتغير المعتاد normal probability plot (NPP)، (3) اختبار جارك- بيرا Jarque- Bera.

المدرج التكراري للبواقي: Histogram of Residuals

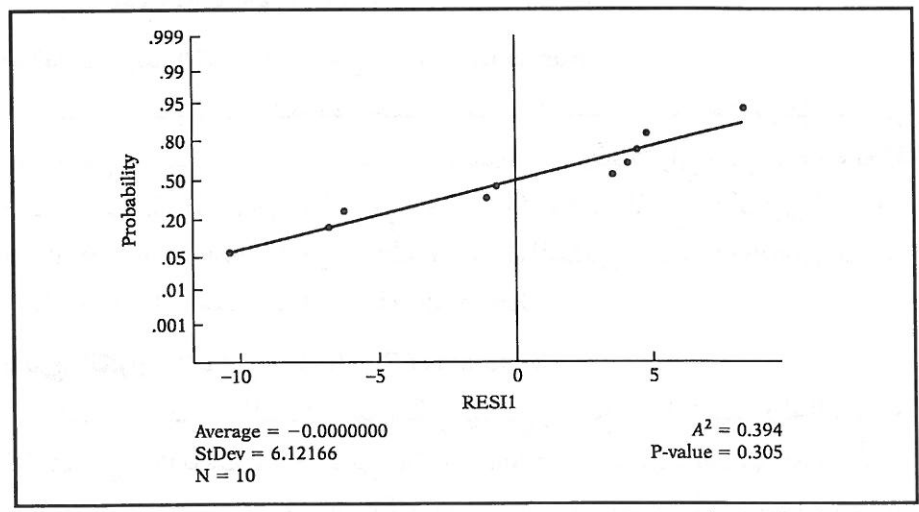
المدرج التكراري للبواقي هو شكل بياني يوضح شكل (نمط) دالة التوزيع الاحتمالي PDF للمتغير العشوائي random variable. حيث يتم تقسيم المحور الأفقي horizontal axis إلى فئات (فترات) تمثل الفئات التي يتم تقسيم المتغير محل الاعتبار

إليها (فمثلاً المتغير الذي يمثل البواقي لتطبيق طريقة المربعات الصغرى OLS residuals) وتقام على كل فئة عمود (مستطيل) ارتفاعه يمثل التكرارات المناظرة للفئة المقام عليها العمود ، حيث تسمى مجموعة الأعمدة الملتصقة التي تمثل التكرارات المناظرة للفئات تسمى بالمدرج التكراري histogram فإذا كان شكل المدرج التكراري يمكن أن يأخذ شكل المنحنى للمتغير المعتاد (الطبيعي) تقريباً. فإنه يمكن القول أن المتغير محل الدراسة يقترب من التوزيع الطبيعي. ومثال ذلك في الفصل 13.5 (انظر شكل 8.5). فهو تدريب جيد لرسم المدرج التكراري للبواقي. وهذه الطريقة تعتبر طريقة بسيطة لاختبار فرض التوزيع الطبيعي للمتغير العشوائي.

رسم الاحتمال الطبيعي : Normal Probability Plot

يناظر رسم دالة كثافة الاحتمال probability density function (PDF) للمتغير الطبيعي (المعتاد) شكلاً آخر يطلق عليه رسم الاحتمال الطبيعي normal probability plot (NPP) حيث يتم وضع قيم المتغير محل الدراسة (وفي هذه الحالة المتغير $\hat{u}_j$) على المحور الأفقي، ووضع قيم y_j المناظرة لها على المحور الرأسي.

فإذا كان المتغير $\hat{u}_j$ مسحوبًا من مجتمع معتاد (طبيعي) normal population، فنجد أن NPP تقترب من الخط المستقيم. فنجد أن NPP لانحدار الدخل والاستهلاك consumption income regression- كما هو موضح في شكل (7.5) الذي تم الحصول عليه باستخدام الحزمة الإحصائية MINIRABS النسخة 13.



شكل (7.5) الحصول على انحدار الدخل والاستهلاك باستخدام الحزمة الإحصائية

وبالتالي إذا كان الشكل التقريبي لـ NPP خطًا مستقيمًا تقريباً ، فهذا يعني أن المتغير محل الاهتمام يتبع التوزيع الطبيعي المعتاد (الطبيعي).

وتستخدم الحزمة الإحصائية MINITAB أيضاً في إجراء اختبار أندرسون- دارلنج لاختبار فرض الطبيعي Anderson- Darling normality test ، المعروف بالإحصاء A^2 . حيث يعتبر الفرض العدمي هو أن المتغير محل الدراسة يتبع التوزيع المعتاد (الطبيعي) والفرض البديل عكس ذلك. وبالنسبة لشكل (5-4) بالنسبة لمثال الدخل والاستهلاك نجد أن قيمة الإحصاء A^2 تساوي 0.305، وهي تعتبر قيمة ملائمة جداً. وبالتالي ، فإنه لايمكن رفض الفرض العدمي القائل أن البواقي ($\hat{u}$) بالنسبة لمثال الدخل والإنفاق يتبع توزيعًا معتاداً. وشكل (7.5)، يوضح أن توقع البواقي يساوي صفراً وانحراف معياري(STDOV) يساوي 6.12166 .

اختبار جارك- بيرا لفرض التوزيع الطبيعي :[20]

Jarque - Bera (JB) test of normality

اختبار JB للتوزيع الطبيعي يعتبر توزيعًا تقاربيًا asymptotic بمعنى أنه اختبار بالنسبة للعينات كبيرة الحجم. ويبنى هذا الاختبار على طريقة المربعات الصغرى للبواقي OLS residuals . ويتطلب هذه الاختبار حساب كل من معامل الالتواء skewness coefficient ومعامل التفرطح kurtosis coefficient (انظر ملحق A) بالنسبة لبواقي المربعات الصغرى ثم استخدم الإحصاء JB حيث :

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right] \qquad (1.12.5)$$

حيث n تساوي حجم العينة، S تساوي معامل الالتواء، K تساوي معامل التفرطح.

حيث إنه بالنسبة للمتغير الذي يتبع التوزيع المعتاد (الطبيعي) نجد أن $S = 0, K = 3$. لذلك نجد أن إحصاء الاختبار JB عندما $S = 0, K = 3$ يجب أن تساوي قيمة الإحصاء JB صفر. وبالتالي عندما JB = 0 أو تقترب من الصفر ، فإن المتغير محل الاعتبار يتبع التوزيع المعتاد أو يقترب منه.

وبالتالي ، تحت افتراض التوزيع الطبيعي للبواقي residuals فنجد أن JB (بالنسبة للعينات الكبيرة) المعطى في العلاقة (1.12.5) يتبع توزيع مربع كا chi-square distribution بدرجات حرية 2. وإذا تم حساب القيمة p المناظرة لـ JB تكون قيمة

(20) See C. M. Jarque and A. K. Bera, “A Test for Normality of Observations and Regression Residuals,” International Statistical Review, vol. 55, 1987, pp. 163-172.

صغيرة عندما تختلف قيمة JB عن الصفر. كذلك نجد أن القيمة p تكون كبيرة نسبياً كلما اقتربت قيمة الإحصاء JB من الصفر، في هذه الحالة لانستطيع رفض الفرض العدمي القائل بأن المتغير محل الاهتمام يتبع التوزيع المعتاد.

وفي مثال **الاستهلاك والدخل**، نجد أن حجم العينة صغير نوعاً ما، لذا عند الحديث بدقة في هذا المثال، يجب عدم استعمال اختبار JB. ولكن رغم ذلك إذا استعملنا اختبار JB في المثال المذكور، نجد أن JB= 0.7769 p= 0.68 عند درجات حرية 2 (من جدول توزيع كا^2) وبنتائج هذا الاختبار، لايمكن رفض الفرض العدمي القائل إن المتغير العشوائي (u_j) يتبع التوزيع المعتاد.

اختبارات أخرى لصلاحية النموذج : Other Tests of Model Adequacy

بالرجوع إلى نموذج CNLRM حيث يتطلب فروضًا أكثر من فرض التوزيع الطبيعي للمتغير العشوائي the error term u_j.

مثال ختامي : A concluding example

إذا عدنا إلى مثال 3-2 حول إنفاق الغذاء food expenditure في الهند India. إذا اعتبرنا البيانات في (2.7.3) وتم تطبيق القانون (1.11.5)، نحن سوف نحصل على معادلة الإنفاق التالية :

$$\begin{aligned}
\widehat{\text{FoodExp}_i} &= 94.2087 + 0.4368\ \text{TotalExp}_i \\
se &= (50.8563) \quad (0.0783) \\
t &= (1.8524) \quad (5.5770) \\
p &= (0.0695) \quad (0.0000)^* \\
r^2 &= 0.3698; \quad df = 53 \\
F_{1,53} &= 31.1034 \quad (p\ \text{value} = 0.0000)^*
\end{aligned} \tag{5.12.2}$$

حيث تشير * إلى أصغر مايمكن.

أولاً : إذا فسرنا الانحدار، كما توقعنا، توجد علاقة طردية positive relationship بين الإنفاق على الطعام والإنفاق الكلي total expenditure. فإذا زاد الإنفاق الكلي بما يساوي روبية rupee عن المتوسط. فإن الإنفاق على الطعام يزيد بمقدار 44 بيسه paise (واحد من 100 من الروبية rupee). وعندما يكون الإنفاق الكلي يساوي صفراً، فإن متوسط الإنفاق على الطعام يساوي 94 روبية. وواضح أن التفسير الميكانيكي لقيمة متوسط الإنفاق على الطعام يساوي 94روبية عندما يكون الإنفاق الكلي يساوي صفراً ليس له دلالة اقتصادية.

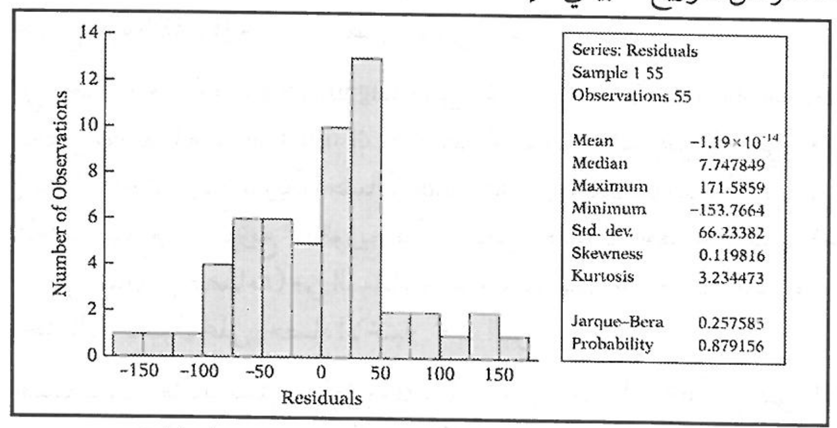
وقيمة r^2= 0.37 تعني أن 37في المائة من الاختلاف في الإنفاق على الطعام ترجع إلى الإنفاق الكلي بدلاً من الدخل. وبافتراض أننا نرغب في اختبار الفرض العدمي أنه

لاتوجد علاقة بين الإنفاق على الطعام والإنفاق الكلي أو بعبارة أخرى اختيار أن 0 =β_2. بما أن قيمة β_2 المقدرة تساوي 0.4368، فإذا كان الفرض العدمي صحيحًا، فما هو احتمال الحصول على القيمة 0.4368؟ تحت صحة الفرض العدمي، من العلاقة (2.12.5)، نحن نجد أن قيمة t تساوي 5.5770، وقيمة p المناظرة لـ t تساوي صفراً. بعبارة أخرى يمكن رفض الفرض العدمي. ولكن بافتراض أن الفرض العدمي 0.5 =β_2 فإنه في هذه الحالة نجد أن :

$$t = \frac{0.4368 - 0.5}{0.0783} = -0.8071$$

فنجد أن احتمال الحصول على 0.807 = $|t|$ أكبر من 20% . لهذا نحن لانرفض الفرض القائل بأن 0.5 =β_2. كذلك نلاحظ أنه تحت صحة الفرض العدمي أن 0 =β_2 فإن F= 31.1034 كما هو واضح في العلاقة (2.12.5) . وتحت نفس الفرض العدمي ، نجد أن t= 5.577، كذلك t^2= 31.1029، وهي تقريباً نفس قيمة F، ومن هنا تتضح العلاقة الوثيقة بين كل من الإحصاء F والإحصاء t (لاحظ : أن عدد درجات الحرية للبسط تساوي واحداً في هذه الحالة).

وباستخدام تقدير البواقي للانحدار، ماهو التوزيع الاحتمالي للحد العشوائي error term؟ والمعلومات موضحة في شكل (8.5). وكما هو موضح في الشكل البواقي للانحدار الإنفاق على الطعام يبدو أنه متماثل التوزيع to be symmetrically distribtued . وتطبيق اختبار JB يوضح أن الإحصاء JB يساوي تقريباً 0.2576 باحتمال يساوي 88% تحت افتراض التوزيع الطبيعي للإحصاء.



شكل (8.5) المدرج التكراري للبواقي من انحدار الإنفاق على الطعام

لذلك نحن لانرفض الفرض العدمي القائل بأن حدود الخطأ error terms تتبع التوزيع الطبيعي. ولكن يجب أن تتذكر أن حجم العينة 55 مشاهدة لايعتبر حجمًا كبيراً بدرجة كافية.

ويترك للقارئ بناء فترات الثقة لمعاملات الانحدار طالما أن NPP تم تحديدها.

13.5 الملخص والاستنتاجات : SUMMARY AND CONCLUSIONS

1 - التقديرات واختبارات الفروض يعتبران الفرعين الرئيسيين في الإحصاء التقليدي classical statistics ففي الفصلين (3 ، 4) تم مناقشة التقديرات. وفي هذا الفصل ، تم تناول اختبارات الفروض.

2 - اختبارات الفروض تجاوب على السؤال التالي : هل يمكن إقرار الفرض العدمي أم لا؟

3 - يوجد أسلوبان متنافيان ومتكاملان للإجابة على السؤال التالي فترات الثقة confidence interval واختبارات المعنوية test of significance .

4 - أسلوب فترة الثقة يهتم بتقدير الفترة. حيث الفترة المقدر هي الفترة أو المدى range الذي يحتوي على القيمة الحقيقية (الفعلية) للمعلمة parameter المطلوب تقديرها وذلك باحتمال معين. وهذا المدى هو مايسمى بفترة الثقة confidence interval وعادة تكتب فترة الثقة وفقاً للاحتمال المحدد ، وبالتالي دائماً فترة الثقة تأخذ شكل النسبة المئوية مثل 90% = $(2 \le \beta_2 \le 5)$. ومما هو جدير بالذكر أنه في حالة وقوع قيمة المعلمة في الفرض العدمي داخل فترة الثقة ، فإنه في هذه الحالة لايمكن رفض الفرض العدمي ، ولكن إذا كانت قيمة المعلمة في الفرض العدمي خارج فترة الثقة ، فإنه يجب رفض الفرض العدمي.

5 - في اختبار المعنوية significance test، نحن نكون إحصاء الاختبار test statistic ثم فحص توزيع المعاينة sampling distribution تحت صحة الفرض العدمي. وعادة إحصاء الاختبار يتبع توزيعًا احتماليًا probability distribution معروفًا مثل التوزيع المعتاد، توزيع t، وتوزيع F، توزيع كا^2. وبمجرد حساب إحصاء الاختبار (على سبيل المثال الإحصاء t) من البيانات يتم تحديد قيمة p. حيث تعطي قيمة p احتمال الحصول على إحصاء الاختبار تحت الفرض العدمي.

6 - عملياً يجب تحديد مستوى المعنوية α بعناية. وهي احتمال الخطأ من النوع الأول type I error مثل 10 ، 5 ، 1 في المائة، كذلك حساب القيمة p.

7 - وبالطبع ، اختبارات الفروض تفترض أن النموذج المختار للتحليل التجريبي empirical analysis مناسب ، بمعنى أنه لايتعارض مع فرض أو أكثر من فروض نموذج الانحدار الخطي الطبيعي التقليدي the classical normal linear regression model .

لذلك من الملائم أن اختبارات النموذج تسبق اختبارات الفروض tests of hypothesis لاكتشاف هل حد الخطأ (المتغير العشوائي) error term تتبع التوزيع الطبيعي أم لا.

في حالة العينات صغيرة الحجم أو المجتمعات المحدودة finite ، فإن ذلك يتطلب إجراء اختبارات t ,F أو اختبارات كا^2 لاختبار فرض أن الخطأ العشوائي يتبع التوزيع الطبيعي.

8 - وإذا كان النموذج ملائمًا عملياً، فإنه يمكن استخدامه في التنبؤ forecasting. ولكن التنبؤ بالقيم المستقبلية للمتغير التابع Y يجب أن يكون في إطار العينة المستخدمة في بناء النموذج فيما عدا ذلك تتزايد أخطاء التنبؤ forecasting errors.

تمارين: EXERCISES

أسئلة : Question

1.5 حدد أي العبارات التالية صحيحة true وأيها خطأ false ، أو غير معينة uncertain أو معينة (محددة) Be precise.

(أ) اختبار t للمعنوية الذي تم مناقشته في هذا الفصل يتطلب أن توزيعات المعاينة sampling distributions للتقديرات $\hat{\beta}_1$ و $\hat{\beta}_2$ تتبع التوزيع الطبيعي.

(ب) عندما المتغير العشوائي (حد الخطأ) في نموذج CLRM لايتبع التوزيع الطبيعي، فإن تقديرات المربعات الصغرى لمعلمات النموذج تظل غير تقديرات غير متحيزة unbiased.

(ج) في نموذج الانحدار إذا كانت قيمة $\hat{\beta}_1 = 0$ فإن تقدير u_i $(\hat{u}_i=)$ مجموعها يختلف عن الصفر.

(د) قيمة p تساوي توقع إحصاء الاختبار test statistic mean.

(هـ) في حالة إذا كان $\hat{\beta}_1 \neq 0$ في نموذج الانحدار ، فإن دائماً مجموع البواقي يساوي صفراً.

(و) في حالة عدم رفض الفرض العدمي، فإنه يكون صحيحًا.

(ز) تساوي قيمة التوقع الشرطي conditional mean والتوقع غير الشرطي unconditional means للمتغير.

(ح) في نموذج الانحدار في متغير مستقل واحد، إذا كان $\beta_2=0$ فإن تقدير β_1 يساوي $\bar{Y}$.

(ط) التباين الشرطي conditional variance var $(Y_i|X_i)=\sigma^2$، والتباين غير الشرطي unconditional variance للمتغير Y حيث $var(Y)=\sigma_Y^2$، يكونان متساويين إذا كانت Y لاتعتمد على X.

2.5 في ضوء جدول تحليل التباين ANOVA في جدول (4.5) بالنسبة لنموذج الانحدار في (2.7.3) واختبارات الفروض، فإنه لاتوجد علاقة بين الإنفاق على الطعام والإنفاق الكلي في الهند.

3.5 باستخدام البيانات المعطاة في جدول (6.2) عن المكسب earnings والتعليم education، أمكن الحصول على النموذج التالي (انظر المعادلة (3.7.3)) :

$$\widehat{\text{Meanwage}}_i = 0.7437 + 0.6416\ \text{Education}_i$$
$$\text{se} = (0.8355) \quad (\quad)$$
$$t = (\quad) \quad (9.6536) \quad r^2 = 0.8944 \quad n = 13$$

(أ) اكتب الأعداد الناقصة بين القوسين ( ).

(ب) مامعنى قيمة $\beta_1=0.6416$.

(جـ) هل أنت ترفض الفرض القائل بأن مستوى التعليم ليس له أثر على الأجر؟ وماهو الاختبار الذي تستخدمه لذلك. ماهي قيمة p المناظرة لإحصاء الاختبارات المستخدمة؟

(د) كون جدول تحليل التباين ANOVA لهذا المثال، ثم اختبر الفرض العدمي القائل بأن معامل الانحدار β_2 تساوي صفراً. ماهو الاختبار المستخدم ولماذا؟

(هـ) افترض أن نموذج الانحدار المعطى أعلاه أن قيمة r^2 غير معطاة. هل يمكن الحصول على قيمة r^2 باستخدام معلومات أخرى من نموذج الانحدار المعطى.

4.5 إذا اعتبرنا أن ρ^2 تشير إلى معامل الارتباط في المجتمع true population coefficient of correlation . إذا فرضنا أننا نرغب في اختبار الفرض العدمي $\rho^2=0$. اشرح كيف يتم إجراء الاختبار (استخدم المعادلة (11.5.3)). انظر أيضاً تمرين 7.5.

5.5 مايعرف بالخط التميزي characteristic line لتحليل الاستثمار الحديث يعتبر خط انحدار regression line يمكن الحصول عليه من النموذج التالي :

$$r_{it} = \alpha_i + \beta_i r_{mt} + u_t$$

حيث :

معدل الفائدة على المأمونية i في الزمن t $= r_{it}$

معدل العائد على حافظة الأوراق السوقية في الزمن t $= r_{mt}$

قيمة الحد العشوائي (المتغير العشوائي) في الزمن t $= u_t$

حيث تسمى β_i بمعامل بيتا beta coefficient للمأمونية رقم (i)، أي مقياس خطر السوق market risk للمأمونية(*).

وعلى أساس 240 معدلاً شهريًا للعائد خلال الفترة 1956–1976 . كل من Fogler and Ganapathy حصلا على الخط التميزي لاسم IBM بعلاقتها بالرقم القياسي للحافظة السوقية the market portifolio index عن طريق جامعة شيكاغو(**).

$$\hat{r}_{it} = 0.7264 + 1.0598 r_{mt} \qquad r^2 = 0.4710$$
$$\text{se} = (0.3001)\,(0.0728) \qquad \text{df} = 238$$
$$F_{1,238} = 211.896$$

(أ) عندما يكون معامل بيتا للمأمونية أكبر من واحد ، فإنه يقال في هذه الحالة المأمونية هاشة (ضعيفة). هل كانت المأمونية لـ IBM هاشة خلال الفترة محل الدراسة؟

(ب) هل β_1 تختلف معنوياً عن الصفر؟ وإذا كانت كذلك ماهو التفسير العملي practical meaning لذلك؟

6.5 يمكن إعادة صياغة المعادلة (5.2.5) على النحو التالي :

$$\Pr\,[\hat{\beta}_2 - t_{\alpha/2}\text{se}\,(\hat{\beta}_2) < \beta_2 < \hat{\beta}_2 + t_{\alpha/2}\text{se}\,(\hat{\beta}_2)] = 1 - \alpha$$

لماذا المتباينة الضعيفة weak inequality (≤) في هذه الحالة يمكن إحلالها بمتباينة قوية strong inequality .

(*) See Haim Levy and Marshall Sarnat, Portfolio and Investment Selection: Theory and Practice, Prentice-Hall International, Englewood Cliffs, N. J., 1984, Chap. 12.

(**) H. Russell Fogler and Sundaram Ganapathy, Financial Econometrics, Prentice Hall, Englewood Cliffs, N.J., 1982, p. 13.

7.5 فيشر R.A. Fisher اشتق توزيع المعاينة لمعامل الارتباط في العينة كما في (13.5.3). وإذا فرضنا أن كلاً من المتغيرين X, Y لهما توزيع احتمالي طبيعي مشترك jointly normally distributed، عندما يتبعان التوزيع الطبيعي الثنائي Bivariate normal distribution (انظر ملحق A4، تمرين 1.4)، وتحت صحة الفرض العدمي أن معامل الارتباط $\rho = 0$، أنه يتضح أن $t = r\sqrt{n-2}/\sqrt{1-r^2}$ متغير يتبع توزيع استيودنت t بدرجات حرية $(n-2)$(*). حيث تعرف قيمة t كما هو موضح بالعلاقة (2.3.5) تحت الفرض العدمي $\beta_2 = 0$. ووفقاً لهذا الفرض العدمي نجد أن $F = t^2$ (انظر الفصل 9.5).

Problems مسائل :

8.5 إذا كانت مخرجات أحد نماذج الانحدار على النحو التالي(**) :

$$\hat{Y}_i = 0.2033 + 0.6560X_t$$
$$\text{se} = (0.0976) \quad (0.1961)$$
$$r^2 = 0.397 \qquad \text{RSS} = 0.0544 \qquad \text{ESS} = 0.0358$$

حيث معدل مشاركة (مساهمة) قوة العمل $Y=$ (LFPR) للمرأة في سنة 1972 .

LFPR للمرأة في سنة 1968 $X=$ حيث حجم العينة المستخلص منها نتائج الانحدار تساوي 19 مدينة في الولايات المتحدة.

(أ) كيف يمكن تفسير نموذج الانحدار أعلاه؟

(ب) اختبر الفرض العدمي H_0 حيث

$$H_0 : \beta_2 = 1$$

ضد الفرض البديل H_1 حيث

$$H_1 : \beta_2 > 1$$

ما هو الاختبار المستخدم في هذه الحالة. ولماذا؟ ماهي الفروض assumptions التي يجب توافرها لإجراء الاختبار المستخدم؟

(*) If ρ is in fact zero, Fisher has shown that r follows the same t distribution provided either X or Y is normally distributed. But if ρ is not equal to zero, both variables must be normally distributed. But if ρ is not equal to zero, both variables must be normally distributed. See R. L. Anderson and T. A. Bancroft, Statistical Theory in Research, McGraw-Hill, New York, 1952, pp. 87–88.

(**) Adapted from Samprit Chatterjee, Ali S. Hadi, and Bertram Price, Regression Analysis by Example, 3d ed., Wiley Interscience, New York, 2000, pp. 46–74.

(جـ) افترض أن معدل مساهمة قوة العمل LFPR في سنة 1968 يساوي 0.58 (أي 58%). وفقاً لنموذج الانحدار أعلاه، ماهو متوسط معدل مساهمة قوة العمل mean LFPR في سنة 1972؟ كون فترة ثقة بدرجة ثقة 95% لمتوسط المتغير التابع Y.

(د) كيف يمكن اختبار أن المتغير العشوائي error term في انحدار المجتمع population regression يتبع التوزيع المعتاد (الطبيعي)؟ ثم وضح الخطوات الضرورية لإجراء الاختبار.

9.5 جدول (5.5) يعطي بيانات عن متوسط راتب المدرس في المدارس العادية (السنوي بالدولار) والمنفق على التلميذ في هذه المدارس (بالدولار) في عام 1985 لـ 50 ولاية ومقاطعة في كولومبيا Columbia.

جدول (5.5) يوضح متوسط الراتب والمنفق على التلميذ (بالدولار) 1985

Observation	Salary	Spending	Observation	Salary	Spending
1	19,583	3346	27	22,795	3366
2	20,263	3114	28	21,570	2920
3	20,325	3554	29	22,080	2980
4	26,800	4642	30	22,250	3731
5	29,470	4669	31	20,940	2853
6	26,610	4888	32	21,800	2533
7	30,678	5710	33	22,934	2729
8	27,170	5536	34	18,443	2305
9	25,853	4168	35	19,538	2642
10	24,500	3547	36	20,460	3124
11	24,274	3159	37	21,419	2752
12	27,170	3621	38	25,160	3429
13	30,168	3782	39	22,482	3947
14	26,525	4247	40	20,969	2509
15	27,360	3982	41	27,224	5440
16	21,690	3568	42	25,892	4042
17	21,974	3155	43	22,644	3402
18	20,816	3059	44	24,640	2829
19	18,095	2967	45	22,341	2297
20	20,939	3285	46	25,610	2932
21	22,644	3914	47	26,015	3705
22	24,624	4517	48	25,788	4123
23	27,186	4349	49	29,132	3608
24	33,990	5020	50	41,480	8349
25	23,382	3594	51	25,845	3766
26	20,627	2821			

Source: Ntionl Eduction Associtio, s reported by Albuquerque Tribune, Nov. 7, 1986.

ولتحديد العلاقة بين الراتب المدفوع للمدرس والمنفق على التلميذ في هذه المدارس يقترح النموذج التالي :

$$Pay_i = \beta_1 + \beta_2 \, Spend_i + u_i$$

حيث تشير pay_i إلى متوسط الراتب المدفوع للمدرس في الولاية (i)، $spend_i$ تشير إلى متوسط المنفق على التلميذ في المدارس العادية في الولاية (i) :

(أ) ارسم البيانات في الجدول السابق مع خط الانحدار.

(ب) من (أ) قدر معلمات النموذج ، ثم أوجد الخطأ المعياري standard errors، كذلك r^2، RSS ، ESS.

(جـ) فسر نموذج الانحدار. فسر النموذج اقتصادياً.

(د) كون فترة الثقة للمعلمة β_2 بدرجة ثقة 95% . هل ترفض الفرض العدمي أن معامل الانحدار (β_2) يساوي 3.0؟

(هـ) إذا فرضنا أن $spend_i = \$ 5000$، قدر متوسط الراتب للمدرس في هذه الحالة. ثم أوجد فترة الثقة للمتوسط الفعلي لراتب مدرس بدرجة ثقة 95% بشرط $spend_i = \$ 5000$.

(و) كيف يمكن اختبار فرض التوزيع الطبيعي للمتغير العشوائي (حد الخطأ) error term ؟ مع توضيح الاختبار (أو الاختبارات) الذي يمكن إجراؤه (أو إجراؤها).

10.5 ارجع إلى تمرين 20.3 وباستخدام جدول ANOVA، اختبر الفرض القائل بأنه لاتوجد علاقة بين الإنتاجية والأجر.

11.5 ارجع إلى تمرين 7.1 :

(أ) ارسم البيانات بحيث يمثل المحور الأفقي المنفق على الإعلانات، والمحور الرأسي التأثير الراجع للإعلان. مانوع العلاقة المشاهدة من البيانات؟

(ب) هل ملائم توفيق نموذج انحدار ثنائي المتغيرات؟ ولماذا نعم؟ أو لماذا لا؟ وفي حالة عدم إمكانية توفيق النموذج، مانوع النموذج المناسب توفيقه؟ وهل لدينا أدوات ضرورية لتوفيق هذه النماذج؟

(جـ) في حالة عدم رسم البيانات- وتم توفيق نموذج انحدار بسيط للبيانات أوجد مخرجات الانحدار. ثم احفظ النتائج لاستخدامها فيما بعد.

12.5 بالرجوع إلى تمرين 1.1 :

(أ) ارسم الرقم القياسي لسعر المستهلك في الولايات المتحدة U.S consumer

price index (CPI) مقابل الرقم القياسي لسعر المستهلك في كندا the Canadian (CPI). ماذا يوضح الرسم.

(ب) بافتراض الرغبة في التنبؤ بالرقم القياسي لسعر المستهلك في الولايات المتحدة على أساس الرقم القياسي لسعر المستهلك في كندا. كون نموذجًا مناسبًا.

(جـ) اختبر الفرض العدمي القائل بعدم وجود علاقة بين الرقم القياسي لسعر المستهلك في الولايات المتحدة والرقم القياسي لسعر المستهلك في كندا. استخدم مستوى المعنوية 5% = α. في حالة رفض الفرض العدمي. هل هذا يعني أن الرقم القياسي لسعر المستهلك في كندا مؤثر في الرقم القياسي في الولايات المتحدة؟ لماذا نعم أو لماذا لا؟

13.5 بالرجوع إلى تمرين 22.3 :

(أ) قدر معلمات نماذج الانحدار ثم أوجد الخطأ المعياري standard errors والمخرجات العادية الأخرى للنموذج.

(ب) اختبر الفرض القائل بأن الأخطاء العشوائية distrubances في نموذجي الانحدار تتبع التوزيع المعتاد (الطبيعي).

(جـ) في انحدار السعر الذهبي the gold price regression، اختبر الفرض القائل أن $\beta_2 = 1$، وتوجد علاقة واحد إلى واحد one- to one relationship بين السعر الذهبي والرقم القياسي لسعر المستهلك (CPI) (بمعنى أن السعر الذهبي يعتبر سياجًا تامًا). ماهي قيمة p لتقدير إحصاء الاختبار test statistic.

(د) كرر (جـ) بالنسبة لانحدار الرقم القياسي NYSE the NYSE index regression. هل الاستثمار في البورصة يعتبر سياجًا تامًا ضد التضخم inflation؟ وماهو الفرض العدمي المستخدم في هذه الحالة؟ وماهي القيمة p؟

(هـ) بين الذهب والبورصة، ماهو الاستثمار الذي يمكن اختباره؟ وماهو الأساس المستخدم في اتخاذ القرار؟

14.5 جدول (6.5) يعطي بيانات GNP، وأربعة تعريفات للمخزون النقدي money stock للولايات المتحدة خلال الفترة 1970–1983 . وبانحدار GNP على التعريفات المختلفة للمخزون النقدي، نحصل على النتائج الموضحة في جدول (7.5).

والتحليل النظري للدخل العادي norminal income (بمعنى GNP العادي) يتحدد بشكل شامل عن طريق التغيرات في الكميات أو المخزون النقدي، بالرغم من عدم وجود إجماع على صحة التعريف للمخزون النقدي. فإذا اعتبرنا النتائج في الجدول التالي، اعتبر الأسئلة التالية:

(أ) ماهو التعريف النقدي الذي يبدو مرتبطًا تماماً بـ GNP العادي؟

(ب) بما أن قيم r^2 كبيرة بشكل عام، فهل هذا يعني أن الاختلاف في تعريف المخزون النقدي لا يعتبر ذا تأثير فعال.

15.5 افترض أن المعادلة المناظرة لمنحنى عدم التحيز indifference curve بين نوعين من البضائع على النحو التالي:

$$X_i Y_i = \beta_1 + \beta_2 X_i$$

كيف يمكن تقدير معلمات النموذج السابق؟ طبق النموذج السابق على البيانات في جدول (5–8) ثم عقب على النتائج.

جدول (6.5) تقدير معلمات نموذج المنحنى لعدم التحيز

GNP AND FOUR MEASURES OF MONEY STOCK

Year	GNP, $ billion	Money stock measure, $ billion			
		M_1	M_2	M_3	L
1970	992.70	216.6	628.2	677.5	816.3
1971	1,077.6	230.8	712.8	776.2	903.1
1972	1,185.9	252.0	805.2	886.0	1,023.0
1973	1,326.4	265.9	861.0	985.0	1,141.7
1974	1,434.2	277.6	908.5	1,070.5	1,249.3
1975	1,549.2	291.2	1,023.3	1,174.2	1,367.9
1976	1,718.0	310.4	1,163.6	1,311.9	1,516.6
1977	1,918.3	335.4	1,286.7	1,472.9	1,704.7
1978	2,163.9	363.1	1,389.1	1,647.1	1,910.6
1979	2,417.8	389.1	1,498.5	1,804.8	2,117.1
1980	2,631.7	414.9	1,632.6	1,990.0	2,326.2
1981	2,957.8	441.9	1,796.6	2,238.2	2,599.8
1982	3,069.3	480.5	1,965.4	2,462.5	2,870.8
1983	3,304.8	525.4	2,196.3	2,710.4	3,183.1

Definitions:

M_1 = currency + demand deposits + travelers checks and other checkable deposits (OCDs)

M_2 = M_1 + overnight RPs and Eurodollars + MMMF (money market mutual fund) balances + MMDAs (money market deposit accounts) + savings and small deposits

M_3 = M_2 + large time deposits + term RPs + Institutional MMMF

L = M_3 + other liquid assets

Source: Economic Report of the President, 1985, GNP data from Table B-1, p. 232; money stock data from Table B-61, p. 303.

جدول (7.5) بيانات GNP وتعريفات للمخزون النقدي

GNP–MONEY STOCK REGRESSIONS, 1970–1983

1)	$\widehat{GNP}_t = -787.4723 + 8.0863\, M_{1t}$ $(77.9664) \quad (0.2197)$	$r^2 = 0.9912$
2)	$\widehat{GNP}_t = -44.0626 + 1.5875\, M_{2t}$ $(61.0134) \quad (0.0448)$	$r^2 = 0.9905$
3)	$\widehat{GNP}_t = 159.1366 + 1.2034\, M_{3t}$ $(42.9882) \quad (0.0262)$	$r^2 = 0.9943$
4)	$\widehat{GNP}_t = 164.2071 + 1.0290\, L_t$ $(44.7658) \quad (0.0234)$	$r^2 = 0.9938$

Note: The figures in parentheses are the estimated standard errors.

جدول 8.5

Consumption of good *X*:	1	2	3	4	5
Consumption of good *Y*:	4	3.5	2.8	1.9	0.8

16.5 في سنة 1986 قدم أحد المتخصصين في الاقتصاد القياسي دراسة عن الرقم القياسي لسعر بيع معطف المطر ذي الحجم الكبير big mac index، لقياس أي العملات الدولية international currencies تكون أصح كمعدل التبادل exchange rate، باستخدام نظرية قيمة القوة الشرائية purchasing power parity (PPP). حيث تتناول قيمة القوة الشرائية PPP وحدة العملة التي تكون قادرة على شراء نفس مجموعة السلع في كل الدول.

مقترحات PPP، ثبتت أنه في المدى الطويل long run، العملات تميل إلى الحركة تجاه قيمة القوة الشرائية لها PPP. حيث استخدم المتخصص Mc Donald's لمعاطف المطر ذات الحجم الكبير كحزمة (أو مجموعة) من السلع، كما هو معطى في جدول (9.5).

اعتبر نموذج الانحدار التالي :

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

حيث Y تساوي معدل التبادل الفعلي actual exchange rate، X تساوي القوة الشرائية المشمولة PPP للدولار.

جدول (9.5) استخدام نظرية قيمة القوة الشرائية (PPP)

	Big Mac prices		Implied PPP* of the dollar	Actual \$ exchange rate April 17, 2001	Under (−)/ over (+) valuation against the dollar, %
	In local currency	In dollars			
United States†	\$2.54	2.54	–	–	–
Argentina	Peso2.50	2.50	0.98	1.00	−2
Australia	A\$3.00	1.52	1.18	1.98	−40
Brazil	Real3.60	1.64	1.42	2.19	−35
Britain	£1.99	2.85	1.28‡	1.43‡	12
Canada	C\$3.33	2.14	1.31	1.56	−16
Chile	Peso1260	2.10	496	601	−17
China	Yuan9.90	1.20	3.90	8.28	−53
Czech Rep	Koruna56.00	1.43	22.0	39.0	−44
Denmark	DKr24.75	2.93	9.74	8.46	15
Euro area	€2.57	2.27	0.99§	0.88§	−11
France	FFr18.5	2.49	7.28	7.44	−2
Germany	DM5.10	2.30	2.01	2.22	−9
Italy	Lire4300	1.96	1693	2195	−23
Spain	Pta395	2.09	156	189	−18
Hong Kong	HK\$10.70	1.37	4.21	7.80	−46
Hungary	Forint399	1.32	157	303	−48
Indonesia	Rupiah14700	1.35	5787	10855	−47
Japan	¥294	2.38	116	124	−6
Malaysia	M\$4.52	1.19	1.78	3.80	−53
Mexico	Peso21.9	2.36	8.62	9.29	−7
New Zealand	NZ\$3.60	1.46	1.42	2.47	−43
Philippines	Peso59.00	1.17	23.2	50.3	−54
Poland	Zloty5.90	1.46	2.32	4.03	−42
Russia	Rouble35.00	1.21	13.8	28.9	−52
Singapore	S\$3.30	1.82	1.30	1.81	−28
South Africa	Rand9.70	1.19	3.82	8.13	−53
South Korea	Won3000	2.27	1181	1325	−11
Sweden	SKr24.0	2.33	9.45	10.28	−8
Switzerland	SFr6.30	3.65	2.48	1.73	44
Taiwan	NT\$70.0	2.13	27.6	32.9	−16
Thailand	Baht55.0	1.21	21.7	45.5	−52

*Purchasing power parity: local price divided by price in the United States.
†Average of New York, Chicago, San Francisco, and Atlanta.
‡Dollars per pound.
§Dollars per euro.
Source: McDonald's; *The Economist*, April 21, 2001.

(أ) في حالة تحديد PPP ماهي قيم β_1 و β_2 المتوقعة؟

(ب) هل نتائج نموذج الانحدار تتوافق مع التوقع في (P)؟ وماهو الاختبار الإحصائي الذي يمكن استخدامه لاختبار افتراضاتك؟

(جـ) هل يستمر المتخصص في إيجاد الرقم القياسي لمعاطف المطر ذي الحجم الكبير؟ لماذا نعم أو لماذا لا؟

17.5 بالرجوع إلى بيانات تمرين 16.2 . افترض أنك ترغب في التنبؤ بدرجة الطالب في math (Y) بناء على درجة الطالبة في math (X) عن طريق نموذج الانحدار :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

(أ) قدر معلمات النموذج $\hat{\beta}_1$ و $\hat{\beta}_2$.

(ب) من تقدير البواقي the estimated residuals، تحقق من توافر فرض التوزيع الطبيعي لـ u_i.

(جـ) اختبر الفرض العدمي $\beta_1 = 1$، في حالة عدم رفض الفرض العدمي هل هذا يعني وجود علاقة واحد إلى واحد بالنسبة للعلاقة بين درجة التلميذ ودرجة التلميذة في math .

(د) كون جدول ANOVA للمشكلة السابقة .

18.5 كرر التمرين السابق مع اعتبار $X|Y$، درجة الطالب ودرجة الطالبة على التوالي في اللغة .

19.5 اعتبر جدول (10.5) المتضمن للبيانات السنوية الرقم القياسي لسعر المستهلك (CPI) والرقم القياسي لسعر الجملة the wholesale price index (WPI)، وأيضاً المسمى بالرقم القياسي لسعر الإنتاج producer price index (PPI) بالنسبة لاقتصاد الولايات المتحدة خلال الفترة 1960–1999 .

(أ) ارسم CPI على المحور الرأسي، WPI على المحور الأفقي ، ثم وضح نوع العلاقة بين الرقمين القياسيين؟ ولماذا؟

جدول (10.5) البيانات السنوية للرقم القياسي لسعر المستهلك CPI

Year	CPI	WPI	Year	CPI	WPI
1960	29.8	31.7	1980	86.3	93.8
1961	30.0	31.6	1981	94.0	98.8
1962	30.4	31.6	1982	97.6	100.5
1963	30.9	31.6	1983	101.3	102.3
1964	31.2	31.7	1984	105.3	103.5
1965	31.8	32.8	1985	109.3	103.6
1966	32.9	33.3	1986	110.5	99.70
1967	33.9	33.7	1987	115.4	104.2
1968	35.5	34.6	1988	120.5	109.0
1969	37.7	36.3	1989	126.1	113.0
1970	39.8	37.1	1990	133.8	118.7
1971	41.1	38.6	1991	137.9	115.9

تابع - جدول (10.5)

1972	42.5	41.1	1992	141.9	117.6
1973	46.2	47.4	1993	145.8	118.6
1974	51.9	57.3	1994	149.7	121.9
1975	55.5	59.7	1995	153.5	125.7
1976	58.2	62.5	1996	158.6	128.8
1977	62.1	66.2	1997	161.3	126.7
1978	67.7	72.7	1998	163.9	122.7
1979	76.7	83.4	1999	168.3	128.0

source: Economic Report of the President, 2000, pp. 373 and 379.

(ب) بافتراض أنك ترغب في التنبؤ ببعض الأرقام القياسية على أساس أرقام قياسية أخرى. ماهو المتغير التابع Y ، وماهو المتغير المفسر (الأساسي) X؟ ولماذا.

نفذ النموذج في (ب)- ثم أوجد مخرجات النموذج. اختبر الفرض أنه توجد علاقة واحد إلى واحد بين الرقمين القياسيين.

من البواقي التي تم الحصول عليها في (جـ)، اختبر الفرض القائل بأن المتغير العشوائي يتبع التوزيع المعتاد (الطبيعي)؟ ثم حدد الاختبار المستخدم.

ملحق A5 APPENDIX

1.A5 التوزيعات الاحتمالية المرتبطة بالتوزيع الطبيعي :

Probability distributions related to the normal distribution

التوزيعات الاحتمالية t, chi- square (X^2), F تتميز ببعض الخصائص التي تم تناولها في ملحق A، حيث يرتبط معظمها بالتوزيع المعتاد. وفي هذا الملحق، سوف نلخص أهم العلاقات بين التوزيع المعتاد والتوزيعات الأخرى المذكورة من بعض النظريات theorems، وبعض الإثباتات proofs ذي الأهمية في هذا المرجع والموجودة في المراجع الإحصائية[1].

نظرية 1.5

إذا فرضنا أن المتغير $Z_1, Z_2, ..., Z_n$ متغيرات عشوائية مستقلة يتبع كل منها التوزيع الطبيعي بتوقع μ_i، وتباين σ_i^2 بمعنى أن

$$Z_i \sim N(\mu_i, \sigma_i^2)$$

(1) For proofs of the various theorems, see Alexander M. Mood, Franklin A. Graybill, and Duane C. Bose, Introduction to the Theory of Statistics, 3d ed., McGraw-Hill, New York, 1974, pp. 239–249.

فإن المجموع Z حيث :

$$Z = \Sigma k_i Z_i$$

حيث k_i مقادير ثابتة غير سالبة. فإن المتغير Z يتبع التوزيع الطبيعي أيضاً بتوقع $(\Sigma k_i \mu_i)$ وتباين $(\Sigma k_i^2 \sigma_i^2)$ أي أن :

$$Z \sim N(\Sigma k_i \mu_i\, , \Sigma k_i^2 \sigma_i^2)$$

وبعبارة أخرى ، فإن التوليفة الخطية linear combination في متغيرات تتبع التوزيع المعتاد (الطبيعي) يمكن إثبات أنها أيضاً متغير يتبع التوزيع المعتاد أيضاً. فعلى سبيل المثال ، Z_1, Z_2 متغيران مستقلان يتبع كل منهما التوزيع المعتاد بحيث $Z_1 \sim N(10, 2)$ ، $Z_2 \sim N(8, 1.5)$.

وبالتالي ، فإن التوليفة الخطية Z حيث :

$$Z = 0.8Z_1 + 0.2Z_2$$

يتبع التوزيع المعتاد أيضاً بتوقع يساوي :

$$0.8(10) + 0.2(8) = 9.6$$

وتباين يساوي :

$$0.64(2) + 0.04(1.5) = 1.34$$

وبالتالي فإن :

$$Z \sim N(9.6, 1.34)$$

نظرية 2.5

إذا فرضنا أن كل متغير من المتغيرات $Z_1, Z_2, ..., Z_n$ متغيرات غير مستقلة يتبع كل منها التوزيع المعتاد بتوقع μ_i، وتباين σ_i^2، فإن Z حيث :

$$Z = \Sigma k_i Z_i$$

حيث k_i مقادير ثابتة، فإن Z تتبع التوزيع المعتاد أيضاً بتوقع يساوي $(\Sigma k_i \mu_i)$ وتباين يساوي :

$$[\ \Sigma k_i^2 \sigma_i^2 \ + 2\ \Sigma k_i k_j \operatorname{cov}(Z_i, Z_j),\ _{i} \neq j\ ]$$

هكذا، إذا فرضنا أن $Z_1 \sim N(6, 2)$ ، $Z_2 \sim N(7, 3)$ حيث $\operatorname{cov}(Z_1, Z_2) = 0.8$، فإن التوليفة الخطية $0.6Z_1 + 0.4Z_2$

تمثل أيضاً متغيراً يتبع التوزيع المعتاد أيضاً بتوقع يساوي :

$$0.6(6) + 0.4(7) = 6.4$$

وتباين يساوي :

$$[\ 0.36(2) + 0.16(3) + 2(0.6)(0.4)(0.8)\ ] = 1.584$$

نظرية 3.5

إذا فرضنا أن المتغيرات $Z_1, Z_2, ..., Z_n$ متغيرات عشوائية مستقلة يتبع كل منها التوزيع المعتاد القياسي ، أي أن :

$$Z_i \sim N(0, 1)$$

فإن ΣZ_i^2 بحيث :

$$\Sigma Z_i^2 = Z_1^2 + Z_2^2 + ... + Z_n^Z$$

تمثل متغيراً يتبع توزيع مربع كا[2] بدرجة حرية n بمعنى :

$$\Sigma Z_i^2 \sim X_n^2$$

وبعبارة أخرى، فإن مجموع مربعات المتغيرات المستقلة التي يتبع كل منها التوزيع المعتاد القياسي تمثل متغيراً يتبع توزيع مربع كا[2] بدرجة حرية تساوي عدد المتغيرات[2].

نظرية 4.5

إذا فرضنا أن $Z_1, Z_2, ..., Z_n$ متغيرات عشوائية مستقلة يتبع كل منها توزيع X^2 بدرجات حرية k_i فإن ΣZ_i حيث :

$$\Sigma Z_i = Z_1 + Z_2 + ... + Z_n$$

تمثل متغيراً يتبع توزيع كا[2] أيضاً بدرجات حرية k حيث :

$$k = \Sigma k_i$$

وتسمى هذه النظرية بخاصية إعادة الإنتاج لـ X^2.

نظرية 5.5

إذا فرضنا أن Z_1 متغير يتبع التوزيع المعتاد القياسي [بمعنى $(Z_1 \sim N(0, 1)$]، والمتغير Z_2 متغير يتبع توزيع مربع كا[2] بدرجات حرية k (بمعنى $Z_2 \sim X^2$) حيث Z_1، Z_2 متغيران مستقلان، فإن المتغير t حيث :

(2) Ibid., p. 243.

$$t=\frac{Z_1}{\sqrt{Z_2}/\sqrt{k}}=\frac{Z_1\sqrt{k}}{\sqrt{Z_2}}=\frac{\text{متغير معتاد قياس}}{\sqrt{\frac{X^2\ \text{متغير يتبع}}{df}}}\sim t_k$$

أي أن المتغير t متغير يتبع توزيع أستيودنت (t) بدرجات حرية k.

ملحوظة : تم مناقشة هذا التوزيع في ملحق A كما هو موضح في الفصل (5). ويلاحظ أن عدد درجات الحرية k ممكن أن تتزايد زيادة غير محدودة بحيث $k\to\infty$ في هذه الحالة ، فإن توزيع أستيودنت t_k عندما $k\to\infty$ يؤول إلى التوزيع المعتاد القياسي[3].

نظرية 6.5

إذا كان المتغيران المستقلان Z_1, Z_2 كل منهما يتبع توزيع مربع كا^2 بدرجات حرية k_1, k_2 على الترتيب ، فإن المتغير F حيث :

$$F=\frac{Z_1/k_1}{Z_2/k_2}\sim F_{k_1,k_2}$$

أي أن المتغير F متغير يتبع توزيع فيشر F بدرجات حرية k_1 (للبسط)، k_2 (للمقام) وبالتالي ، فإن نظرية 5–6 توضح أن خارج قسمة متغيرين كل منهما يتبع توزيع مربع كا^2 يساوي متغير يتبع توزيع F.

نظرية 7.5

مربع المتغير الذي يتبع توزيع t بدرجات حرية k يمثل متغيراً يتبع توزيع F بدرجات حريه $k_1=1 \mid k_2=k$[4]. أو بعبارة أخرى :

$$F_{1,k}=t_k^2$$

ومثال ذلك :

$$F_{1,4}=t_4^2 \quad \text{or} \quad F_{1,23}=t_{23}^2$$

نظرية 8.5

عندما تتزايد عدد درجات الحرية للمقام للمتغير العشوائي الذي يتبع توزيع F، فإن المتغير F في هذه الحالة يقترب من توزيع مربع كا^2 بمعنى :

$$m\,F_{m,n}=\chi_m^2 \qquad \text{as } n\to\infty$$

(3) For proof, see Henri Theil, Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, N.J., 1978, pp. 237–245.

(4) For proof, see Eqs. (5.3.2) and (5.9.1).

نظرية 9.5

عندما يتزايد عدد درجات الحرية تتزايد كا^2 في المتغير يتبع توزيع كا^2 فإنه يمكن تقريب الجزر التربيعي لمتغير كا^2 بمتغير معتاد قياسي، أو بعبارة أخرى :

$$Z = \left\{\sqrt{2x^2} - \sqrt{2k-1}\right\} \sim N(0,\ 1)$$

حيث k تساوي عدد درجات الحرية.

2.A5 اشتقاق المعادلة (2.3.5) : Derivation of equation (2.3.5)

اعتبر

$$Z_1 = \frac{\hat{\beta}_2 - \beta_2}{\text{se}\,(\hat{\beta}_2)} = \frac{(\hat{\beta}_2 - \beta_2)\sqrt{x_i^2}}{\sigma} \qquad (1)$$

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} \qquad (2)$$

حيث التباين σ^2 معلوم، فإن المتغير Z_1 يتبع التوزيع المعتاد القياسي، أي أن $Z_1 \sim N(0, 1)$ (لماذا)؟، والمتغير Z_2 يتبع توزيع كا^2 بدرجات حرية $(n-2)$[5] :

$$Z^2 \sim X^2_{(n-1)}$$

ويمكن توضيح أن Z_1، Z_2 متغيرين مستقلين[6].

وباستخدام نظرية 5.5 نجد أن المتغير t حيث :

$$t = \frac{Z_1\sqrt{n-2}}{\sqrt{Z_2}} \qquad (3)$$

يتبع توزيع t بدرجات حرية $(n-2)$. بالتعويض بـ (1)، (2) في (3) نحصل على المعادلة (2.3.5).

3.A5 اشتقاق المعادلة (1.9.5) : Derivation of equation (1.9.5)

المعادلة رقم (1) توضح أن المتغير Z يتبع التوزيع المعتاد القياسي، أي أن $Z_1 \sim N(0, 1)$. لذلك باستخدام النظرية (3.5) نجد أن المتغير التالي :

(5) For proof, see Robert V. Hogg and Allen T. Craig, Introduction to Mathematical Statistics, 2d ed., Macmillan, New York, 1965, p. 144.

(6) For proof, see J. Johnston, Econometric Methods, McGraw-Hill, 3d ed., New York, 1984, pp. 181–182. (Knowledge of matrix algebra is required to follow the proof.)

$$Z_1^2 = \frac{(\hat{\beta}_2 - \beta_2)^2 \sum x_i^2}{\sigma^2}$$

متغير يتبع توزيع مربع كا^2 بدرجة حرية تساوي واحد. ونلاحظ من الفصل (1.A5) أن :

$$Z_2 = (n-2)\frac{\hat{\sigma}^2}{\sigma^2} = \frac{\sum \hat{u}_i^2}{\sigma^2}$$

متغير يتبع توزيع مربع كا^2 بدرجات حرية تساوي (n–2). فضلاً عن ذلك نجد من الفصل (3.4) أن المتغير Z_3 مستقل عن المتغير Z_1. من النظرية 6.5 نجد أن :

$$F = \frac{Z_1^2/1}{Z_2/(n-2)} = \frac{(\hat{\beta}_2 - \beta_2)^2(\sum x_i^2)}{\sum \hat{u}_i^2/(n-2)}$$

متغير يتبع توزيع F بدرجات حرية 1 في البسط، (n–2) في المقام. وتحت الفرض العدمي H_0: $\beta_2 = 0$ نجد أن النسبة F كما هي بالمعادلة (1.9.5).

4.A5 اشتقاق المعادلات من (2.10.5)، (6.10.5): Derivation of equations

تباين توقع التنبؤ: Variance of mean prediction

بافتراض $X_i = X_0$، نجد أن توقع التنبؤ الحقيقي يساوي $E(Y_0|X_0)$ على النحو التالي :

$$E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \quad (1)$$

وبتقدير (1) من :

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \quad (2)$$

وبأخذ التوقع لطرفي العلاقة (2) نجد أن :

$$E(\hat{Y}_0) = E(\hat{\beta}_1) + E(\hat{\beta}_2) X_0$$
$$= \beta_1 + \beta_2 X_0$$

حيث β_1 و β_2 تقديران غير متحيزين unbiased estimators لذلك :

$$E(\hat{Y}_0) = E(Y_0 \mid X_0) = \beta_1 + \beta_2 X_0 \quad (3)$$

وكذلك نجد أن $\hat{Y}_0$ تقدير غير متحيز أيضاً لـ $E(Y_0|X_0)$.

والآن، سوف نستخدم الخاصية التالية :

$$\text{Var}(a+b) = \text{var}(a) + \text{var}(b) + 2\,\text{cov}(a,b)$$

سوف نحصل على :

$$\text{var}(\hat{Y}_0) = \text{var}(\hat{\beta}_1) + \text{var}(\hat{\beta}_2)X_0^2 + 2\,\text{cov}(\hat{\beta}_1\hat{\beta}_2)X_0 \tag{4}$$

وباستخدام صياغات التباين والتغير covariance لكل من $\hat{\beta}_1$ و $\hat{\beta}_2$ في (9.3.3) ، (3.3.3) ، (1.3.3) نحصل على :

$$\text{var}(\hat{Y}_0) = \sigma^2\left[\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \quad = (2.10.5)$$

تباين القيمة المتنبأ بها : Variance of individual prediction

إذا فرضنا أننا نرغب في التنبؤ بالوحدة Y بشرط $X = X_0$، حيث :

$$Y_0 = \beta_1 + \beta_2 X_0 + u_0 \tag{5}$$

كذلك :

$$\hat{Y}_0 = \hat{\beta}_1 + \hat{\beta}_2 X_0 \tag{6}$$

فإن خطأ التنبؤ $(Y_0 - Y_0)$ على النحو التالي :

$$\begin{aligned} Y_0 - \hat{Y}_0 &= \beta_1 + \beta_2 X_0 + u_0 - (\hat{\beta}_1 + \hat{\beta}_2 X_0) \\ &= (\beta_1 - \hat{\beta}_1) + (\beta_2 - \hat{\beta}_2)X_0 + u_0 \end{aligned} \tag{7}$$

لذلك فإن :

$$\begin{aligned} E(Y_0 - \hat{Y}_0) &= E(\beta_1 - \hat{\beta}_1) + E(\beta_2 - \hat{\beta}_2)X_0 - E(u_0) \\ &= 0 \end{aligned}$$

حيث كل من $\hat{\beta}_1$ و $\hat{\beta}_2$ تقديران غير متحيزين، X_0 قيمة محددة، $E(u_0) = 0$ وفقاً للفرض.

وبتربيع طرفي المعادلة (7) ثم أخذ توقع الطرفين ، فإننا نحصل على :

$$\text{var}\left(Y_0 - \hat{Y}_0\right) = \text{var}\left(\hat{\beta}_1\right) + X_0^2\,\text{var}\left(\hat{\beta}_2\right) + 2X_0\,\text{cov}\left(\beta_1, \beta_2\right) + \text{var}\left(u_0\right)$$

وباستخدام صياغات التباين والتغير لكل من $\hat{\beta}_1$ و $\hat{\beta}_2$ مع مراعاة أن $\text{var}(u_0) = \sigma^2$ نجد أن :

$$\text{var}(Y_0 - \hat{Y}_0) = \sigma^2\left[1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum x_i^2}\right] \quad = (6.10.5)$$

الفصل السادس

توسيع نطاق نماذج الانحدار الخطية ثنائية المتغيرات[*]

EXTENSIONS OF THE TWO-VARIABLE LINEAR REGRESSION MODEL[*]

بعض المفاهيم الأساسية في تحليل الانحدار الخطي، يمكن استعراضها بسهولة من خلال نموذج الانحدار الخطي ثنائي المتغيرات. والذي سبق وتناولناه.

أولاً: استعرضنا حالة نموذج الانحدار المار بنقطة الأصل، أي الحالة التي لا يكون في النموذج أي جزء ثابت مقطوع من المحور الأصلي، β_1. **ثانياً**: انتقلنا إلى السؤال عن وحدات القياس، أي كيف يتم قياس المتغيرات X و Y وما إذا كان تغيير وحدات القياس سيؤثر على نتائج الانحدار. **أخيراً**: تناولنا شكل دالة نموذج الانحدار الخطي. وحتى الآن، فإن حديثنا مقتصر على النماذج الخطية في المعلمات والمتغيرات معًا. ولكن تذكر أن نظرية الانحدار التي استعرضناها في الفصول السابقة، تتطلب أن يكون النموذج خطي في المعلمات فقط، بغض النظر عن أنه خطيًا في المتغيرات أم لا. عند التعامل مع نماذج خطية في المعالم، ولكن ليست بالضرورة خطية في المتغيرات، سنستعرض في هذا الفصل بعض المشكلات العملية المهمة التي سيعاني منها النموذج ثنائي المتغيرات.

وبمجرد أن تتضح هذه الفكرة، فإنه من السهل توسيع نطاق التطبيق ليشمل نماذج الانحدار المتعدد، كما سنرى في الفصلين 7 و 8.

1.6 الانحدار المار بنقطة الأصل:

REGRESSION THROUGH THE ORIGIN

هناك بعض الحالات يفترض فيها أن PRF ثنائي المتغيرات يأخذ الشكل التالي:

$$Y_i = \beta_2 X_i + u_i \qquad (1.1.6)$$

(*) الكتاب من ص 211 إلى ص 714 ترجمة أ. م. د. هند عبد الغفار عودة.

في هذا النموذج الجزء الثابت المقطوع من المحور الأصلي غير موجود، أو يساوي الصفر. ومن هنا جاءت تسميته **النموذج المار بنقطة الأصل**.

للتوضيح اعتبر نموذج سعر أصول رأس المال (CAPM) لنظرية السندات التجارية الحديثة، والتي يمكن تمثيلها وفقًا لشكل المخاطر الأولية كالتالي[1]:

$$(ER_i - r_f) = \beta_i (ER_m - r_f) \qquad (2.1.6)$$

حيث ER_i = معدل العائد المتوقع على المدخر i.

ER_m = معدل العائد المتوقع على سندات السوق التجارية والتي تمثل، مثلاً، بـ S&P 500 «المؤشر المركب للأسهم».

r_f = معدل العائد الخالي من المخاطر، مثلاً، عائد 90 يومًا على كشف الحساب.

β_i = معامل Beta، مقياس للمخاطر المنظمة، أي المخاطر التي لا يمكن تجنبها من خلال توظيف الأموال. ويعتبر أيضًا مقياسًا لمدى تغير معدل العائد على المدخر i مع تغييرات السوق. A $\beta_i > 1$ يعني أنه مخالف للحماية أو الضمان أما $\beta_1 < 1$ فيعني أنه متفق ومدافع عن الحماية والأمان (لاحظ التالي: لا تربط خطأ بين β_i وبين معامل الميل الخاص بالنموذج ثنائي المتغيرات، β_2).

إذا كان سوق رأس المال يعمل بكفاءة، فإن CAPM يفترض أن المخاطر الأولية المتوقعة للسند i $(=ER_i - r_f)$ تساوي معامل β للسند مضروب في المخاطرة الأولية المتوقعة للسوق $(=ER_m - r_f)$. إذا تحقق CAPM فإن لدينا وضعًا ما موضح في الشكل (1.6). الخط البياني الموجود في الشكل يُعرف باسم **خط سندات السوق (SML)**.

لأغراض تطبيقه، تتم كتابة (2.1.6) كالتالي:

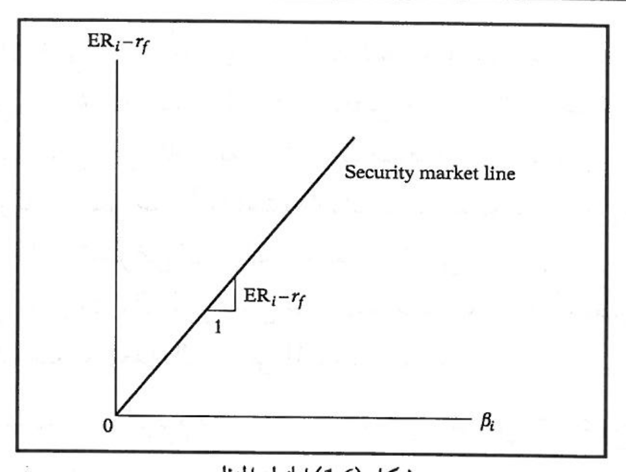
$$R_i - r_f = \beta_i (R_m - r_f) + u_i \qquad (3.1.6)$$

أو

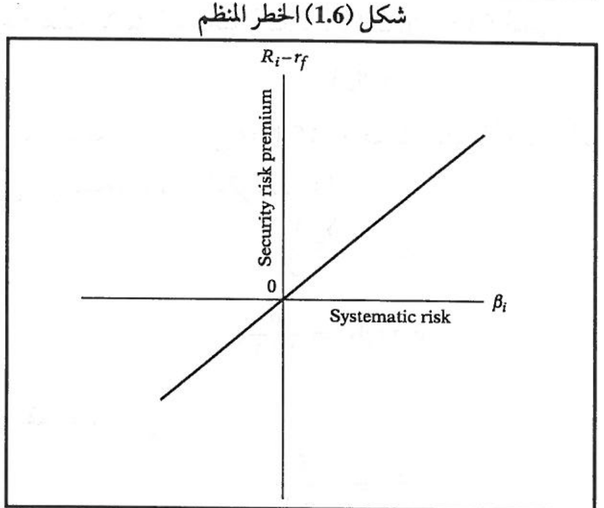
$$Ri - r_f = \alpha_i + \beta_i (R_m - r_f) + u_i \qquad (4.1.6)$$

---

(1) انظر Haim Levy and marshall Sarnat, Portfolio and Investment Selection: Theory and Practice, Prentice-Hall International, Englewood Cliffs, N.J., 1984, Chap. 14.



شكل (1.6) الخطر المنظم



شكل (2.6) : نموذج السوق وفقًا لنظرية السندات التجارية (بافتراض $\alpha_i = 0$)

النموذج الأخير معروف باسم **نموذج السوق**[2]. إذا تحقق CAPM فإن α_i يتوقع أن تساوي الصفر (انظر شكل 2.6).

وبشكل عابر، لاحظ أنه في (4.1.6)، المتغير التابع Y يساوي $R_i - r_f$، والمتغير المفسر X هو β_i، معامل التطاير، وليس $R_m - r_f$. وبالتالي لنقوم بعمل انحدار (4.1.6)، لابد أولاً من تقدير β_i، والتي نحصل عليها عادة من **الخط الوصفي**، كما سبق وفصلناه في تمرين 5.5 (لمزيد من التفاصيل، انظر تمرين 8.28).

(2) انظر، على سبيل المثال، Diana R. Harrington, Modern Portfolio Theory and the Capital Asset Pricing Model: A User's Guide, Prentice Hall, Englewood Cliffs, N.J., 1983, p. 71.

من المثال الحالي، نرى أنه أحيانًا تتطلب النظرية غياب الجزء الثابت المقطوع من المحور الصادي عن نموذج الانحدار. هناك بعض الحالات التي يكون فيها النموذج الذي لا يشتمل على جزء ثابت يعتبر نموذجًا منطقيًا، مثل فرض الدخل لـ Milton Friedman، والذي ينص على أن الاستهلاك الدائم يتناسب مع الدخل الدائم، نظرية تحليل التكلفة التي تفترض أن متغير التكلفة الخاص بالانتاج يتناسب مع الناتج، وأيضًا بعض نظريات المال التي تنص على أن معدل تغير الأسعار (أي معدل التضخم) يتناسب مع معدل التغيير في المعروض من المال.

كيف يمكنك تقدير نموذج مثل (1.1.6)، وما هي المشاكل التي تتوقع وجودها؟

للإجابة عن هذه الأسئلة، دعنا نكتب SRF لـ (1.1.6) كالتالي:

$$Y_i = \hat{\beta}_2 X_i + \hat{u}_i \qquad (5.1.6)$$

والآن بتطبيق طريقة OLS على (5.1.6)، نحصل على المعادلة التالية لـ$\hat{\beta}_2$ وتباينيا (الإثبات معطي في الملحق A6، الفقرة 1.A6)

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (6.1.6)$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum X_i^2} \qquad (7.1.6)$$

حيث σ^2 مقدرة كالتالي:

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-1} \qquad (8.1.6)$$

من المثير مقارنة هذه المعادلات مع نظيرها الذي نحصل عليها عندما يحتوي النموذج على جزء ثابت مقطوع من المحور الصادي:

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \qquad (3.1.6)$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad (1.3.3)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \qquad (5.3.3)$$

الفرق بين المجموعتين من المعادلات، يجب أن يكون واضحًا، وهو كالتالي: في النموذج الذي لا يشتمل على جزء ثابت، نستخدم مجموع مربعات خام،

وحواصل ضرب. ولكن في النموذج الذي يوجد فيه جزء ثابت، نستخدم مجموع مربعات معدل (بالنسبة للوسط الحسابي) وحواصل ضرب. ثانيًا: درجات الحرية المستخدمة لحساب σ^2 تساوي $n-1$ في الحالة الأولى و $n-2$ في الحالة الثانية. (لماذا؟)

على الرغم من أن النماذج التي لا تشتمل على جزء ثابت تكون مناسبة في العديد من الحالات، فإن هناك بعض الخصائص التي يجب ملاحظتها عند التعامل مع مثل هذه النماذج. أولاً، $\Sigma\hat{u}_i$ الذي يساوي عادة الصفر في النموذج الذي يوجد فيه جزء ثابت (النموذج التقليدي) قد لا يساوي الصفر في النموذج الذي لا يحتوي على جزء ثابت. باختصار، $\Sigma\hat{u}_i$ قد لا تساوي الصفر في حالة النموذج المار بنقطة الأصل. ثانيًا، r^2، معامل التحديد الذي سبق واستعرضنا. في الفصل 3، والذي دائمًا يكون غير سالب في النموذج التقليدي، قد يكون في بعض الحالات سالبًا، كما في حالة النماذج التي لا تحتوي على جزء ثابت.

وبالتالي فـ r^2 المحسوبة تقليديًا قد لا تناسب نماذج الانحدار المارة بنقطة الأصل[3].

r^2 لنماذج الانحدار المارة بنقطة الأصل: r^2 for Regression-through-Origin Model

كما سبق وذكرنا، وكما سنجد في ملحق A6، الفقرة 10.A6، r^2 التقليدية المعطاة في الفصل 3 قد لا تكون مناسبة للنماذج التي لا تشتمل على جزء ثابت. ولكن يمكن حساب القيمة المعروفة باسم **r^2 الخام** لمثل هذه النماذج، والمعرفة كالتالي:

$$\text{raw } r^2 = \frac{\left(\sum X_i Y_i\right)^2}{\sum X_i^2 \sum Y_i^2} \qquad (9.1.6)$$

لاحظ أن: مجموع مربعات خام (أي غير مصحح للوسط الحسابي)، وأيضًا حواصل الضرب على الرغم من أن r^2 الخام مستوفية العلاقة $0 < r^2 < 1$، لا يمكن مقارنتها مباشرةً مع قيمة r^2 التقليدية. ولهذا السبب، فإن بعض الكتاب لا يعتمدون على قيمة r^2 لنماذج الانحدار التي لا تحتوي على جزء ثابت.

نظرًا لهذه الخصائص المرتبطة بمثل هذا النموذج، لابد للباحث أن يتعامل بحذر شديد مع نماذج الانحدار التي لا تشتمل على جزء ثابت. وبالتالي إذا لم تكن هناك أسباب مسبقة مهمة لاستخدام نماذج الانحدار المارة بنقطة الأصل، فإنه من الأفضل استخدام نماذج الانحدار التي تحتوي على الجزء الثابت المقطوع من المحور الصادي.

---

(3) المزيد من التفاصيل، انظر، Dennis J. Aigner, basic Econometrics, Prentice Hall Englewood Cliffs, N.J., 1971. pp. 85–88.

ويوجد لهذا منفعة مزدوجة. أولاً: إذا كان النموذج يحتوي على الجزء الثابت، ولكن وجدنا أنه إحصائيًا غير معنوي (أي أنه إحصائيًا يساوي الصفر) فإنه وفقًا لأي أغراض تطبيقية فكأن لدينا نموذجًا مارًا بنقطة الأصل[4]. ثانيًا والأكثر أهمية، إذا كان بالفعل هناك جزء ثابت، ولكننا استخدمنا نموذج مارًا بنقطة الأصل، فإننا نقع في خطأ توصيفي، مما يجعلنا نخالف الفرض 9 من نموذج الانحدار الخطي الكلاسيكي.

مثال توضيحي:

الخط المميز لنظرية محافظة الأوراق المالية : The characteristic line of portfolio theory

جدول (1.6) يعطي بيانات خاصة بمعدل العائد السنوي (%) على تمويل مستقبلي، وهو تمويل هدفه الأساسي تعظيم مكسب رأس المال. ويقاس ذلك في سوق السندات التجارية بمؤشر Fisher خلال الفترة 1971 إلى 1980.

في تمرين (5.5) استعرضنا الخط المميز لتحليل الاستثمار، والذي يمكن كتابته كالتالي:

$$Y_i = \alpha_i + \beta_i X_i + u_i \qquad (10.1.6)$$

حيث Y_i = معدل العائد السنوي (%) على التمويل المستقبلي

X_i = معدل العائد السنوي (%) على السندات التجارية السوقية

β_i = معامل الميل، والمعروف باسم معامل Beta في نظرية السندات التجارية

و

α_i = الجزء الثابت المقطوع من المحور الصادي

في الأدبيات لا توجد أي مرجعية لقيمة α_i. بعض النتائج العملية أثبتت أن هذه القيم تكون موجبة، ولها معنوية إحصائية. وبعض النتائج الأخرى أثبتت العكس، أي أن ليس لها معنوية إحصائية وبالتالي لا تختلف فعليًا عن الصفر. وفقًا لهذه الحالة الأخيرة، يمكن كتابة النموذج كالتالي:

$$Y_i = \beta_i X_i + u_i \qquad (11.1.6)$$

أي انحدار مار بنقطة الأصل.

(4) Henri Theil أوضح أنه إذا كان الجزء الثابت غير موجود فعلاً، فإن معامل الميل يمكن تقديره بدقة أعلى أكثر من الحالة التي يوجد فيها الجزء الثابت: انظر في Introduction to Econometrics, Prentice Hall, Englewood Cliffs, N.J., 1978, p76. انظر أيضًا في المثال الرقمي المعطى لاحقًا.

جدول (1.6)

معدلات العائد السنوية على التمويل المستقبلي ومؤشر Fisher (سوق السندات التجارية) 1971 - 1980

Year	Return on Afuture Fund, % Y	Return on Fisher Index, % X
1971	67.5	19.5
1972	19.2	8.5
1973	−35.2	−29.3
1974	−42.0	−26.5
1975	63.7	61.9
1976	19.3	45.5
1977	3.6	9.5
1978	20.0	14.0
1979	40.3	35.3
1980	37.5	31.0

المصـدر: Haim Levy and Marshall Sarnat, Portfolio and Investment selection: Theory and Practice, Prentice-Hall International, Englewood Cliffs, N. J., 1984, pp. 730 and 738. These data were obtanined by the authors from Weisenberg Investment Service, Investment Companies, 1981 edition.

اذا قررت أن تستخدم النموذج (11.1.6)، ستحصل على نتائج الانحدار التالية:

$$\hat{Y}_i = 1.0899\ X_i$$
$$(0.1916) \qquad \text{raw } r^2 = 0.7825 \qquad \textbf{(6.1.12)} \qquad \textbf{(12.1.6)}$$
$$t = (5.6884)$$

ويتضح من النتائج السابقة أن β_i معنويًا أكبر من الصفر، وتفسير ذلك أن لكل 1% زيادة في معدل عائد السوق، سيزداد معدل عائد التمويل المستقبلي في المتوسط بحوالي 1.09%.

كـيـف يمكننا أن نتـأكـد من أن النمـوذج (11.1.6) وليس (10.1.6) هو النمـوذج المناسب، خصوصًا وأنه لا يوجد تصور في الواقع، بأن α_i تساوي الصفر؟ يمكننا التأكد من ذلك بعمل انحدار (10.1.6). باستخدام البيانات المعطاة في جدول (1.6)، فنحصل على النتائج التالية:

$$\hat{Y}_i = 1.2797 + 1.0691 X_i$$
$$(7.6886) \quad (0.2383) \qquad \textbf{(6.1.13)} \qquad \textbf{(13.1.6)}$$
$$t = (0.1664) \quad (4.4860) \qquad r^2 = 0.7155$$

لاحظ أن: قيمة r^2 الموجودة في (12.1.6) و (13.1.6) لا يمكن مقارنتهما مباشرةً. من هذه النتائج نرفض الفرض القائل بأن الجزء الثابت الحقيقي يساوي الصفر، مما يفسر استخدام (1.1.6)، أي الانحدار المار بنقطة الأصل.

عمومـاً، لاحظ أنه لا يوجد فرق جوهري بين نتائج (12.1.6) و(13.1.6)، إلا أن الأخطاء القياسية لـ β أقل قليلاً في حالة نموذج الانحدار المار بنقطة الأصل، وذلك يدعم مقولة theil المعطاة في الهامش 4، والتي تنص على أنه إذا كانت α_i في الحقيقة تساوي

الصفر، فإن معامل الميل يقاس بدقة أكبر: باستخدام البيانات المعطاة في جدول (1.6) ووفقاً لنتائج الانحدار يمكن للقارئ أن يثبت أن 95% فترة ثقة لمعامل الميل في الانحدار المار بنقطة الأصل هي (0.6566, 1.5232) أما وفقًا للنموذج (13.1.6) فإنها (0.5195, 1.6186)، أي أن الفترة الأولى أضيق من الفترة الأخيرة.

2.6 المقياس ووحدات القياس:

SCALING AND UNITS OF MEASUREMENT

لتوضيح فكرة هذه الفقرة، دعنا نستخدم البيانات المعطاة في جدول (2.6)، والخاصة بإجمالي الاستثمار الخاص المحلي (GPDI) والناتج الإجمالي المحلي (GDP) في الولايات المتحدة بالبليون والمليون من دولارات 1992.

جدول (2.6) إجمالي الاستثمار الخاص المحلي و GDP، الولايات المتحدة، 1988–1997.

Observation	GPDIBL	GPDIM	GDPB	GDPM
1988	828.2000	828200.0	5865.200	5865200
1989	863.5000	863500.0	6062.000	6062000
1990	815.0000	815000.0	6136.300	6136300
1991	738.1000	738100.0	6079.400	6079400
1992	790.4000	790400.0	6244.400	6244400
1993	863.6000	863600.0	6389.600	6389600
1994	975.7000	975700.0	6610.700	6610700
1995	996.1000	996100.0	6761.600	6761600
1996	1084.1000	1084100.0	6994.800	6994800
1997	1206.4000	1206400.0	7269.800	7269800

ملحوظة: GPDIBL = الاستثمار الخاص المحلي، بليون من دولارات 1992.
GPDIM = الاستثمار الخاص المحلي، مليون من دولارات 1992.
GDPB = الناتج المحلي الكلي، بليون من دولارات 1992.
GDPM = الناتج المحلي الكلي، مليون من دورات 1992.

المصدر: Economic Report of the President, 1999, Table B-2, p. 328

افترض أن أحد الباحثين عند قيامه بعمل انحدار لـ GPDI على GDP استخدم البليون دولار، وباحث آخر استخدم البيانات في صورة مليون دولار. هل ستختلف نتائج الانحدارين؟ وإذا حدث ذلك فأي من الانحدارين يجب أن نستخدم؟ بإختصار هل وحدات القياس المستخدمة لقياس المتغيرات المنحدرة والمتغير المنحدر عليه تؤثر على نتائج الانحدار؟ وإذا حدث ذلك، ما هو الأسلوب الأفضل اتباعه لاختيار وحدات القياس المناسبة لتحليل الانحدار؟ للإجابة على هذه الأسئلة، دعنا نتبع الطريقة المنظمة التالية:

اعتبر أن:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i \qquad (1.2.6)$$

حيث Y = GPDI و X = GDP . عرّف :

$$Y_i^* = w_1 Y_i \quad (2.2.6)$$

$$X_i^* = w_2 X_i \quad (3.2.6)$$

حيث w_1 و w_2 ثوابت، تسمى عوامل الأوزان، w_1 و w_2 قد يتساويان وقد لا يتساويان.

من (2.2.6) و(3.2.6) يتضح أن Y_i^* وX_i^* هي Y_i و X_i معدلة الوزن. وبالتالي إذا كان Y_i و X_i مقاسين بالبليون دولار، وأراد الباحث التعبير عنهما في صورة المليون دولار، فإن $Y_i^* = 1000\ Y_i$ و $X_i^* = 1000\ X_i$ ، وهنا يكون $w_1 = w_2 = 1000$. والآن دعنا نعتبر الانحدار المستخدم فيه المتغيرين Y_i^* و X_i^* كالتالي :

$$Y_i^* = \hat{\beta}_1^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \quad (4.2.6)$$

حيث $Y_i^* = w_1 Y_i$ ، $X_i^* = w_2 X_i$ و $\hat{u}_i^* = w_1 \hat{u}_i$ (لماذا؟). نريد الآن أن نكتشف العلاقات الموجودة بين الأزواج التالية :

1. $\hat{\beta}_1$ و $\hat{\beta}_1^*$
2. $\hat{\beta}_2$ و $\hat{\beta}_2^*$
3. $\text{var}(\hat{\beta}_1)$ و $\text{var}(\hat{\beta}_1^*)$
4. $\text{var}(\hat{\beta}_2)$ و $\text{var}(\hat{\beta}_2^*)$
5. $\hat{\sigma}^2$ و $\hat{\sigma}^{*2}$
6. r_{xy}^2 و $r_{x^*y^*}^2$

من نظرية المربعات الصغرى، نعلم التالي (انظر الفصل 3)

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad (5.2.6)$$

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2} \quad (6.2.6)$$

$$\text{var}(\hat{\beta}_1) = \frac{\sum X_i^2}{n \sum x_i^2} \cdot \sigma^2 \quad (7.2.6)$$

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \quad (8.2.6)$$

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} \quad (9.2.6)$$

بتطبيق طريقة OLS على (4.2.6)، نحصل بالمثل على التالي :

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \quad (10.2.6)$$

$$\hat{\beta}_2^* = \frac{\sum x_i^* y_i^*}{\sum x_i^{*2}} \quad (11.2.6)$$

$$\operatorname{var}(\hat{\beta}_1^*) = \frac{\sum X_i^{*2}}{n \sum x_i^{*2}} \cdot \sigma^{*2} \quad (12.2.6)$$

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sigma^{*2}}{\sum x_i^{*2}} \quad (13.2.6)$$

$$\hat{\sigma}^{*2} = \frac{\sum \hat{u}_i^{*2}}{(n-2)} \quad (14.2.6)$$

من هذه النتائج من السهل إيجاد العلاقة بين كل زوج من مقدرات العاملات. كل ما نحتاج إليه هو استخدام العلاقات التالية :

$Y_i^* = w_1 Y_i$ (أو $y_i^* = w_1 y_i$) و $X_i^* = w_2 X_i$ (أو $X_i^* = w_2 x_i$) و $\hat{u}_i^* = w_1 \hat{u}_i$ ،

و $\bar{Y}^* = w_1 \bar{Y}$ و $\bar{X}^* = w_2 \bar{X}$ ، ووفقًا لهذه التعاريف، يمكن للقارئ أن يثبت بسهولة التالي :

$$\hat{\beta}_2^* = \left(\frac{w_1}{w_2}\right) \hat{\beta}_2 \quad (15.2.6)$$

$$\hat{\beta}_1^* = w_1 \hat{\beta}_1 \quad (16.2.6)$$

$$\hat{\sigma}^{*2} = w_1^2 \hat{\sigma}^2 \quad (17.2.6)$$

$$\operatorname{var}(\hat{\beta}_1^*) = w_1^2 \operatorname{var}(\hat{\beta}_1) \quad (18.2.6)$$

$$\operatorname{var}(\hat{\beta}_2^*) = \left(\frac{w_1}{w_2}\right)^2 \operatorname{var}(\hat{\beta}_2) \quad (19.2.6)$$

$$r_{xy}^2 = r_{x^* y^*}^2 \quad (20.2.6)$$

من النتائج السابقة ، يتضح أن نتائج الانحدار بناء على وحدة قياس معينة يمكن استنتاجها بناء على وحدة قياس أخرى، وفقًا لعوامل الأوزان w's. في الواقع يختار الباحث وحدة القياس الأكثر منطقية وسهولة بالنسبة له ، مع الوضع في الاعتبار ، أن التعبير عن الأرقام بعدد ما من الأصفار، وفقًا للمليون أو البليون من الدولارات.

وفقًا للنتائج المعطاة من (15.2.6) وحتى (20.2.6) يمكن استنتاج بعض الحالات الخاصة. مثلاً، إذا كان $w_1 = w_2$، أي عوامل الأوزان متساوية ، فإن معامل الميل وأخطاءه القياسية لن تتأثر باستخدام (Y_i, X_i) أو (X_i^* و Y_i^*). عمومًا، فإن الجزء

الثابت وخطاءه القياسي كلاً منهما مضروب في العامل w_i. ولكن إذا لم يتغير مقياس X (أي $w_2 = 1$) ، ولكن مقياس y تغير بالعامل w_1 ، فإن الميل والجزء الثابت وأخطاءهما القياسية سيكونون جميعًا مضروبين في العامل w_1. وأخيراً إذا لم يتميز مقياس Y (أي $w_1 = 1$) ولكن مقياس X تغير بالعامل w_2 ، وبالتالي فإن معامل الميل وأخطاءه القياسية سيكون مضروبًا في العامل ($1/w_2$) ، أما معامل الجزء الثابت وخطاؤه القياسي فلن يتأثرا.

يجب ملاحظة- عمومًا- أن التحويل من (Y, X) إلى (Y^*, X^*) لا تؤثر على خصائص مقدرات OLS السابق مناقشتها في الفصول السابقة.

مثال رقمي : العلاقة بين GPDI و GDP ، الولايات المتحدة 1988 – 1997

لتوضيح النتائج النظرية السابق عرضها، دعنا نعود إلى البيانات المعطاة في جدول (2.6) ونختبر هذه النتائج (الأرقام المعطاة بين الأقواس تمثل الأخطاء القياسية المقدرة). في النتائج التالية لكل من GPDI و GDP مقاسين بالبليون دولار:

$$\widehat{GPDI}_t = -1026.498 + 0.3016\ GDP_t$$
$$se = (257.5874) \quad (0.0399) \qquad r^2 = 0.8772 \qquad (21.2.6)$$

أما النتائج التالية ، فإن كلاً من GPDI و GDP مقاسان بالمليون دولار:

$$\widehat{GPDI}_t = -1{,}026{,}498 + 0.3016\ GDP_t$$
$$se = (257{,}587.4) \quad (0.0399) \qquad r^2 = 0.8772 \qquad (22.2.6)$$

لاحظ أن الجزء الثابت المقطوع من المحور الصادي وخطائه القياسي يساويان 1000 مرة نظيرهما في الانحدار (21.2.6) (لاحظ أن $w_1 = 1000$ حتى نتحول من البليون إلى المليون دولار). ولكن معامل الميل وخطأه القياسي لم يتغيرا وهذا يتماشى مع الاستنتاجات النظرية السابق عرضها.

الآن دع GPDI بالبليون دولار و GDP بالمليون دولار.

$$\widehat{GPDI}_t = -1026.498 + 0.000301\ GDP_t$$
$$se = (257.5874) \quad (0.0000399) \qquad r^2 = 0.8772 \qquad (23.2.6)$$

كما هو متوقع، معامل الميل وخطاؤه القياسي يساويان 1/1000 من قيمتهما في (21.2.6)، فقد تغير مقياس X، أو GDP، فقط.

والآن دع DPDI بالمليون دولار وGDP بالبليون دولار:

$$\widehat{GPDI}_t = -1{,}026{,}498 + 301.5826\ GDP_t$$
$$se = (257{,}587.4) \quad (39.89989) \qquad r^2 = 0.8772 \qquad (24.2.6)$$

لاحظ مجدداً أن كلاً من الجزء الثابت والميل وأخطائهما القياسية يساويان 1000

مرة مضروب في قيمهم المناظرة في (21.2.6) ، وهذا أيضًا يتفق مع النتائج النظرية السابق عرضها.

لاحظ أنه في كل نتائج الانحدار السابقة ، فإن قيمة r^2 تظل كما هي دون تغيير ، وهذا لا يعتبر شيئاً مفاجئاً ، حيث إن قيمة r^2 لا تتغير مع تغير وحدة القياس ، فهي قيمة خالصة ورقم بدون أبعاد.

ملاحظة خاصة بتفسير النتائج : A Word about Interpretation

بما أن معامل الميل β_2 هو ببساطة معدل متغير ، فإنه يقاس وفقًا لوحدات النسبة :

$$\frac{\text{وحدة المتغير التابع}}{\text{وحدة المتغير المفسر}}$$

وبالتالي في انحدار (21.2.6) يفسر معامل الميل المساوي لـ 0.3016 فإذا تغير الـ GDP بوحدة واحدة ، أي بليون دولار ، فإن GPDI يزداد في المتوسط بـ 0.3016 بليون دولار . في انحدار (23.2.6) ، فإن كل وحدة تغير في الـ GDP أي لكل 1 مليون دولار ، فإن GPDI يزداد في المتوسط بـ 0.000302 بليون دولار . النتيجتان بالطبع متساويتان تمامًا في مدى تأثير GDP على GPDI ، فهي ببساطة نفس النتيجة ولكن معبر عنها بوحدات قياس مختلفة.

3.6 الانحدار وفقًا لمتغيرات قياسية:

REGRESSION ON STANDARDIZED VARIABLES

رأينا في الفقرة السابقة ، أن وحدات قياس المتغير المنحدر عليه ، والمتغير (أو المتغيرات) المنحدرة تؤثر على تفسير نتائج معاملات الانحدار . يمكن تجنب ذلك إذا استخدمنا المتغير المنحدر والمنحدر عليه في صورته القياسية . ويقال إن المتغير في صورته القياسية إذا طرح من قيمة وسطه الحسابي ثم قسمة حاصل الطرح على الانحراف المعياري لهذا المتغير . أي انحدار Y على X ، تستخدم الشكل القياسي التالي :

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_Y} \qquad (1.3.6)$$

$$X_i^* = \frac{X_i - \bar{X}}{S_X} \qquad (2.3.6)$$

حيث $\bar{Y}$ = متوسط العينة لـ Y ، S_Y = الانحراف المعياري للعينة لـ Y ، $\bar{X}$ =متوسط العينة لـ X ، S_X = الانحراف المعياري للعينة لـ X ، المتغيران Y_i^* و X_i^* يسميان متغيرين قياسين.

خاصية مهمة متعلقة بالمتغير القياسي ، هو أن وسطه الحسابي يساوي الصفر وانحرافه المعياري دائمًا يساوي 1 (للإثبات انظر ملحق A6 ، الفقرة 2.A6)

كنتيجة لذلك، لم يعد من المهم معرفة وحدة قياس المتغير المنحدر أو المنحدر عليه. وبالتالي بدلاً من القيام بالانحدار:

$$Y_i = \beta_1 + \beta_3 X_i + u_i \qquad (3.3.6)$$

فإننا نقوم بعمل الانحدار باستخدام المتغيرات القياسية:

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + u_i^* \qquad (4.3.6)$$

$$= \beta_2^* X_i^* + u_i^* \qquad (5.3.6)$$

ويمكن بسهولة إثبات أن الجزء الثابت دائمًا يساوي الصفر في حالة الانحدار الذي يشمل متغيرات منحدرة ومتغيراً منحدراً عليه في الصورة القياسية[5]. معاملات الانحدار الخاصة بالمتغيرات القياسية، والتي نرمز لها بـ β_1^* و β_2^* ، تعرف باسم معاملات بيتا[6]. لاحظ أن انحدار (5.3.6) هو انحدار مار بنقطة الأصل.

كيف يمكن أن نفسر معاملات بيتا السابقة؟ التفسير هو أنه إذا زاد المتغير المنحدر (القياسي) بوحدة انحراف معياري واحدة، فإنه في المتوسط يزداد المتغير المنحدر عليه (القياسي) بـ β_2^* وحدة. وذلك على خلاف النموذج التقليدي (3.3.6)، حيث إن التأثير يقاس في صورة الوحدات الأصلية المقاسة بها Y و X وليس في وحدات الانحراف المعياري.

لتوضيح الفرق بين (3.3.6) و(5.3.6)، دعنا نعود إلى مثال GPDI وGDP السابق مناقشته في الفقرة السابقة. نتائج (21.2.6) السابق عرضها معادة مرة أخرى هنا للتوضيح كالتالي:

$$\begin{aligned} \widehat{\text{GPDI}}_t &= -1026.498 \quad + \quad 0.3016\ \text{GDP}_t \\ \text{se} &= \quad (257.5874) \quad (0.0399) \qquad r^2 = 0.8872 \end{aligned} \qquad (6.3.6)$$

حيث GPDI وGDP مقاسان بالبليون دولار.

(5) نذكر أنه من المعادلة (7.1.3) الجزء الثابت = متوسط قيمة المتغير التابع − الميل × متوسط المتغير المنحدر. ولكن في حالات المتغيرات القياسية القيم المتوسطة للمتغير التابع والمتغير المنحدر عليه تساوي الصفر. وبالتالي الجزء الثابت يساوي الصفر أيضًا.

(6) لا تخلط بين معاملات بيتا المذكورة هنا ومعاملات البيتا الموجودة في نظرية الاستثمار.

النتائج الخاصة بـ (5.3.6)، وعلى اعتبار أن المتغيرات ذات النجمة هي المتغيرات القياسية أصبحت كالتالي:

$$\widehat{\text{GPDI}}_t^* = 0.9387\ \text{GDP}_t^* \qquad (7.3.6)$$
$$\text{se} = (0.1149)$$

نحن نعرف كيف يمكننا تفسير الجزء الثابت من (6.3.6) فهو كالتالي: إذا زاد GDP بدولار واحد، فإن GPDI في المتوسط يزداد بحوالي 30 سنتًا. ماذا عن (7.3.6)؟ هنا يكون التفسير كالتالي: إذا زاد GDP (القياسي) بمقدار واحد انحراف معياري، فإن GPDI يزداد في المتوسط بمقدار 0.94 من الانحراف المعياري.

ما هي الميزة التي يمتاز بها نموذج الانحدار القياسي عن النموذج التقليدي؟ هذه الميزة تتضح أكثر إذا كان النموذج يشتمل على أكثر من متغير منحدر واحد. وهذا الموضوع ستتم مناقشته بالتفصيل في الفصل 7. فعندما يتم التعبير عن كل المتغيرات المنحدرة في صورتها القياسية، فإننا بذلك نجعل لها جميعًا أساسًا واحدًا وبالتالي يمكن مقارنتها مباشرة. فإننا بذلك نجعل لها جميعًا أساسًا واحدًا. وبالتالي يمكن مقارنتها مباشرة. فإذا كان معامل متغير منحدر قياسي أكبر من معامل متغير منحدر قياسي أخرى في النموذج، فإن ذلك يعني أن هذا المتغير الأول له دور أكبر في تفسير المتغير المنحدر عليه من المتغير الأخير.

بمعنى آخر، تستطيع هنا أن تستخدم معاملات بيتا كمقياس للتأثير النسبي، أو مدى القوة النسبية للمتغيرات المنحدرة. وسوف نتناول المزيد من التفاصيل الخاصة بهذا الموضوع في الفصلين التاليين.

وقبل أن نترك هذا الموضوع، لابد أن نلاحظ نقطتين مهمتين. **أولاً**: بالنسبة للانحدار القياسي (7.3.6) لم تكن قيمة r^2 معطاة، حيث إن هذا الانحدار مار بنقطة الأصل، وبالتالي r^2 التقليدية لا يمكن تطبيقها كما سبق وأشرنا في الفقرة 1.6. **ثانيًا**: هناك علاقة مثيرة للانتباه بين معاملات β في النموذج التقليدي ومعاملات البيتا السابق ذكرها. فالعلاقة بينهما كالتالي:

$$\hat{\beta}_2^* = \hat{\beta}_2\left(\frac{S_x}{S_y}\right) \qquad (8.3.6)$$

حيث S_x = الانحراف المعياري للعينة الخاص بالمتغير المنحدر X و S_y = الانحراف المعياري للعينة والخاص بالمتغير المنحدر عليه. وبالتالي يمكن استخدام قاعدة المقص

بين β ومعاملات البيتا إذا كنا نعلم الانحراف المعياري للمتغير المنحدر والمنحدر عليه. سنرى في الفصل التالي، أن هذه العلاقة متحققة أيضًا في حالة الانحدار المتعدد. ومتروك للقارئ كتمرين أن يثبت العلاقة (8.3.6) لمثالنا التوضيحي الحالي.

4.6 الأشكال الدالية لنماذج الانحدار:
FUNCTIONAL FORMS OF REGRESSION MODELS

كما سبق وذكرنا في الفصل 2، فإننا نركز بشكل رئيس على النماذج الخطية في المعلمات، وقد تكون خطية أوغير خطية في المتغيرات. في الفقرة التالية، سنتطرق إلى بعض نماذج الانحدار المستخدمة كثيراً، والتي قد تكون غير خطية في المتغيرات، ولكن خطية في المعالم، أو من الممكن استخدام تحويلة مناسبة لجعل هذه النماذج خطية في المتغيرات. وبالتحديد سنناقش نماذج الانحدار التالية:

1 – النموذج الخطي – اللوغاريتمي.

2 – النماذج شبه اللوغاريتمية.

3 – نماذج المقلوب.

4 – نماذج مقلوب اللوغاريتم.

سنناقش خواص هذه النماذج، والحالات المناسبة لاستخدامها، وإمكانية تقديرها. وكل النموذج سيرافقه أمثلة توضيحية مناسبة له.

5.6 كيفية تقدير المرونة: النموذج الخطي - اللوغاريتمي؟
HOW TO MEASURE ELASTICITY: THE LOG-LINEAR MODEL

اعتبر النموذج التالي، والمعروف باسم نموذج الانحدار الأسي:

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \qquad (1.5.6)$$

والذي يمكن التعبير عنه في الصورة المماثلة التالية[7]:

$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i \qquad (2.5.6)$$

(7) لاحظ الخصائص التالية للوغاريتمات: (1) $\ln (AB) = \ln A + \ln B$ ، (2) $\ln (\frac{A}{B}) = \ln A - \ln B$ ، (3) $\ln (A^k) = k \ln A$ ، بافتراض أن A ، B موجبات و k ثابت.

حيث $\ln n$ = اللوغاريتم الطبيعي (أي اللوغاريتم للأساس e، حيث $e = 2.718$)[8].

يمكن كتابة (2.5.6) كالتالي:

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \qquad (3.5.6)$$

حيث $\alpha = \ln \beta_1$، هذا النموذج خطي في المعالم α وβ_2، خطي في لوغاريتم المتغيرات X وY ويمكن تقدير المعالم بانحدار OLS. ويسبب هذه الخطية فإن مثل هذه النماذج تسمى نماذج خطية لوغاريتمية أو لوغاريتمية مزدوجة. إذا استوفت كل فروض نموذج الانحدار الخطي الكلاسيكي، معامل (3.5.6) يمكن تقديرها باستخدام طريقة OLS كالتالي:

$$Y_i^* = \alpha + \beta_2 X_i^* = u_i \qquad (4.5.6)$$

حيث $Y_i^* = \ln Y_i$ و $X_i^* = \ln X_i$. مقدرات OLS $\hat{\beta}_2$ و$\hat{\alpha}$ التي سنحصل عليها هي أفضل مقدرات خطية غير متحيزة للمعامل α و β_2 على الترتيب. هناك خاصية مرتبطة بالنموذج الخطي اللوغاريتمي، والتي تجعله يستخدم كثيراً في المجال التطبيقي. وتتمثل في ان معامل الميل β_2 يستخدم كمقياس لمرونة المتغير Y بالنسبة للـ X، أي أنه يمثل نسبة التغير في Y بالنسبة بالنسبة لتغيير معلوم في X[9].

---

(8) في الواقع ممكن أن نستخدم الأساس، وفيه الأساس يساوي 10. العلاقة بين الأساس الطبيعي والمعتاد كالتالي: $\ln_e X = 2.3026 \log_{10} X$. ومن المعتاد كتابة الأساس الطبيعي في صورة ln والأساس المعتاد في صورة Log بدون كتابة الأساس e و 10 بشكل صريح.

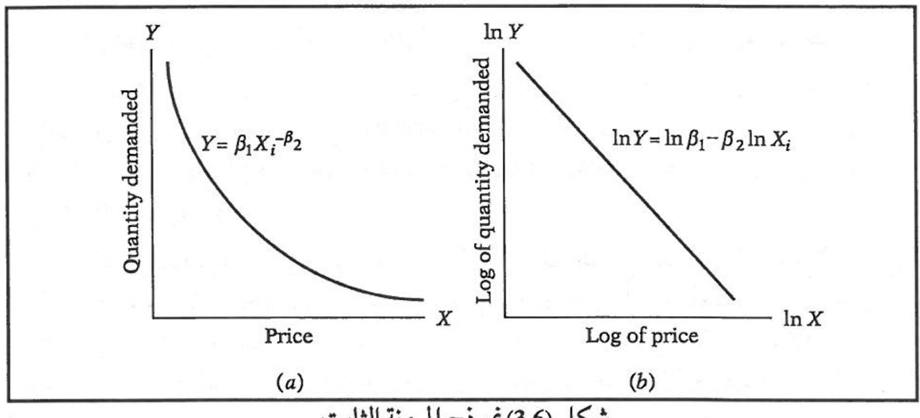
(9) معامل الميل، في الحساب، يعرف كالتالي $[(\frac{dY}{dX})(\frac{X}{Y})] = (\frac{dY}{Y})(\frac{dX}{X})$. القارئ الذي لديه بعض المعلومات الخاصة بالتفاضل سيعلم بسهولة ان β_2 في الحقيقة هي معامل المرونة.

ملاحظة فنية: القارئ العارف بالحساب والتفاضل سيلاحظ أن $d(\ln X)/dX = 1/X$ أو $d(\ln X) = dX/X$، أي أنه بالنسبة للتغير الصغير (لاحظ معامل التفاضل d)، التغير في $\ln X$ سيساوي التغير النسبي في X. في الواقع التطبيقي، إذا كان التغير في X صغيراً، فإنه يمكن كتابة العلاقة كالتالي: التغير في Ln ≑ التغير النسبي في X حيث ≑ تعني تقريباً. وبالتالي بالنسبة للتغير البسيط فإن التغير النسبي في

$$(\ln X_t - \ln X_{t-1}) \doteq (X_t - X_{t-1}) / X_{t-1} = X$$

عمومًا على القارئ أن يأخذ في اعتباره المصطلحات التالية: (1) التغير المطلق، (2) التغير النسبي و(3) تغير النسبة أو معدل النمو النسبي. وبالتالي فإن $(X_t - X_{t-1})$ تمثل تغيراً مطلقًا، $(X_t - X_{t-1}) / X_{t-1} = X_t / X_{t-1} - 1$ يسمى تغيراً نسبيًا $100\ [(X_t - X_{t-1}) / X_{t-1}]$ تسمى معدل النمو تمثل على الترتيب القيمة الحالية والسابقة للمتغير X.

وبالتالي إذا كانت Y تمثل الكمية المطلوبة من سلعة ما، و X يمثل السعر، فإن β_2 تقيس مرونة سعر الطلب، وهذا المعامل يعتبر من المعامل الاقتصادية المهمة. إذا كانت العلاقة بين الكمية المطلوبة والسعر كما في الشكل (a3.6)، التمويلة اللوغاريتمية الثنائية كما موضحة في الشكل (b3.6) ستعطي تقدير لمرونة السعر $(-\beta_2)$.



شكل (3.6) نموذج المرونة الثابت

هناك خاصيتان ممكن ملاحظتهما على النموذج الخطي - اللوغاريتمي: النموذج يفترض أن معامل المرونة بين Y وX، β_2، يظل ثابتًا (لماذا؟) ولذلك سمى النموذج باسم بديل، وهو نموذج المرونة الثابتة[10].

بعبارة أخرى، كما يوضح شكل (b3.6) التغيير في $\ln Y$ بالنسبة لتغيير الوحدة في $\ln X$ (أي المرونة β_2) يظل كما هو بغض النظر عن قيمة $\ln X$ التي نقيس عندها المرونة خاصية أخرى لهذا النموذج متعلقة بأن $\hat{\alpha}$ و $\hat{\beta}_2$ مقدرات غير متحيزة لـ α و β_2 أما β_1 (معامل النموذج الأصلي) عندما تم تقديره بـ$\hat{\beta}_1$ = معكوس اللوغاريتم $(\hat{\alpha})$ فإنه في حد ذاته مقدار متحيز.

في العديد من المشاكل التطبيقية، يكون الجزء الثابت المقطوع من المحور الصادي ثانوي الأهمية، وبالتالي يكون ليس مهماً الحصول على مقدر غير متحيز[11].

(10) نموذج المرونة الثابتة سيعطي تغيراً ثابتًا لإجمالي الربح وفقًا لتغير نسبي معطى في السعر بغض النظر عن السعر المطلق. يمكن للقارئ أن يقارن هذه النتيجة مع شروط المرونة التي تتطلبها دالة الطلب الخطية البسيطة $Y_i = \beta_1 + \beta_2 X_i + u_i$. عمومًا، الدالة الخطية البسيطة تعطي مقدار تغير ثابت للكمية بالنسبة للتغير في السعر. قارن ذلك مع النموذج الخطي اللوغاريتمي وما يتطلبه لكل تغير معطى بالدولار في السعر.

(11) بالنسبة لطبيعة التحيز وما الذي يمكن فعله حيال هذا التحيز. انظر Arthur S. Goldberger, Topics in Regression Analysis, Macmillan, New York, 1978, p. 120.

في النموذج ثنائي المتغيرات، أبسط طريقة لمعرفة ما إذا كان النموذج الخطي اللوغاريتمي مناسب البيانات أم لا، هو أن تقوم برسم شكل انتشار لـ $\ln Y_i$ ضد $\ln X_i$ وترى ما إذا كان شكل الانتشار يأخذ تقريباً خطًا مستقيمًا أم لا كما في شكل (b3.6).

مثال توضيحي:

الإنفاق على السلع المعمرة وعلاقته بنفقات الاستهلاك الإجمالية الشخصية

جدول (3.6) يعطي بيانات عن إجمالي نفقات الاستهلاك الشخصية (PCEXP)، الإنفاق على السلع المعمرة (EXPDUR) والإنفاق على السلع غير المعمرة (EXPNONDUR)، والإنفاق على الخدمات (EXPSERVICES)، هذه القيم كلها مقاسة في 1992 بالبليون دولار.[12]

افترض أننا نرغب في معرفة مرونة الإنفاق على السلع المعمرة بالنسبة لإجمالي إنفاق الاستهلاك الشخصي. برسم لوغاريتم الإنفاق على السلع المعمرة ضد لوغاريتم إجمالي نفقات الاستهلاك الشخصية، نرى أن العلاقة بين هذين المتغيرين علاقة خطية. وبالتالي النموذج اللوغاريتمي المزدوج قد يكون مناسبًا للبيانات. نتائج الانحدار التالي:

$$\begin{aligned} \widehat{\ln \text{EXDUR}_t} &= -9.6971 + 1.9056 \ln \text{PCEX}_t \\ se &= (0.4341) \quad (0.0514) \\ t &= (-22.3370)^* \quad (37.0962)^* \quad r^2 = 0.9849 \end{aligned} \tag{5.5.6}$$

حيث * تعنى أن قيمة P-value صغيرة للغاية.

من هذه النتائج، مرونة EXPDUR بالنسبة لـ PCEX تساوي تقريباً 1.90، مما يعني أنه إذا زاد إجمالي النفقات الشخصية بـ 1% فإنه في المتوسط يزداد الإنفاق على السلع المعمرة بحوالي 1.9%. وبالتالي الإنفاق على السلع المعمرة يتأثر بشكل كبير بالتغير في نفقات الاستهلاك الشخصية. ولهذا السبب يهتم صانع السلع المعمرة بالتغيرات التي تحدث في الدخل الشخصي، والإنفاق الاستهلاكي الشخصي. في تمرين (17.6) و (18.6) يُطلب من القارئ أن يقوم بنفس الخطوات السابقة، ولكن بالنسبة للسلع غير المعمرة والإنفاق على الخدمات.

جدول (3.6) إجمالي النفقات الشخصية والمستويات المختلفة

Observation	EXPSERVICES	EXPDUR	EXPNONDUR	PCEXP
1993-I	2445.3	504.0	1337.5	4286.8
1993-II	2455.9	519.3	1347.8	4322.8
1993-III	2480.0	529.9	1356.8	4366.6
1993-IV	2494.4	542.1	1361.8	4398.0
1994-I	2510.9	550.7	1378.4	4439.4
1994-II	2531.4	558.8	1385.5	4472.2
1994-III	2543.8	561.7	1393.2	4498.2
1994-IV	2555.9	576.6	1402.5	4534.1

(12) تشمل السلع المعمرة السيارات وقطع الغيار والأثاث والمعدات المنزلية. وتشمل السلع غير المعمرة الغذاء والكساء والبنزين والزيت وزيت الوقود والفحم؛ وتشمل الخدمات الإسكان والكهرباء والغاز والنقل، والرعاية الطبية.

تابع –جدول (3.6) إجمالي النفقات الشخصية والمستويات المختلفة

Observation	EXPSERVICES	EXPDUR	EXPNONDUR	PCEXP
1995-I	2570.4	575.2	1410.4	4555.3
1995-II	2594.8	583.5	1415.9	4593.6
1995-III	2610.3	595.3	1418.5	4623.4
1995-IV	2622.9	602.4	1425.6	4650.0
1996-I	2648.5	611.0	1433.5	4692.1
1996-II	2668.4	629.5	1450.4	4746.6
1996-III	2688.1	626.5	1454.7	4768.3
1996-IV	2701.7	637.5	1465.1	4802.6
1997-I	2722.1	656.3	1477.9	4853.4
1997-II	2743.6	653.8	1477.1	4872.7
1997-III	2775.4	679.6	1495.7	4947.0
1997-IV	2804.8	648.8	1494.3	4981.0
1998-I	2829.3	710.3	1521.2	5055.1
1998-II	2866.8	729.4	1540.9	5130.2
1998-III	2904.8	733.7	1549.1	5181.8

لاحظ أن: EXPSERVICES = الإنفاق على الخدمات، بليون دولار 1992.
EXPDUR = الإنفاق على السلع المعمرة، بليون دولار 1992.
EXPNONDUR = الإنفاق على السلع غير المعمرة، بليون دولار 1992.
PCEXP = نفقات الاستهلاك الإجمالية الشخصية، بليون دولار 1992.

المصدر: Economic Report of the Prsident, 1999, Table B-17, p. 347

6.6 النماذج شبه اللوغاريتمية: نماذج LOG–LIN و LIN–LOG SEMILOG MODELS: LOG–LIN AND LIN–LOG MODELS

كيف تقيس معدل النمو: نموذج الـ Log–Lin؟ How to Measure the Growth Rate: The Log–Lin Model

غالبًا ما يهتم كل من الاقتصاديين، رجال الأعمال، ورجال الدولة بمعرفة معدل نمو عدد من المتغيرات الاقتصادية مثل السكان، GNP المعروض من المال، العمالة، الإنتاجية، والعجز التجاري.

افترض أنك تريد معرفة معدل نمو نفقات الاستهلاك الشخصي على الخدمات للبيانات المعطاة في جدول (3.6). اعتبر أن Y_t هي النفقات الحقيقية على الخدمات عند الزمن t و Y_0 قيمة مبدئية للنفقات على الخدمات (مثل، القيمة عند نهاية 1992-IV). استرجع الآن معادلة الفائدة المركبة الذي سبق ودرسناها جميعًا في المواد الدراسية الأساسية في الاقتصاد.

$$Y_t = Y_0(1 + r)^t \qquad (1.6.6)$$

حيث r معدل النمو المركب للـ Y (أي خلال الزمن). بإدخال اللوغاريتم الطبيعي على (1.6.6) نحصل على التالي:

$$\ln Y_t = \ln Y_0' + t\ \ln(1+r) \qquad (2.6.6)$$

والآن دع

$$\beta_1 = \ln Y_0 \qquad (3.6.6)$$

$$\beta_2 = \ln(1+r) \qquad (4.6.6)$$

يمكن الآن كتابة (2.6.6) كالتالي:

$$\ln Y_t = \beta_1 + \beta_2 t \qquad (5.6.6)$$

وبإضافة حد الخطأ إلى (5.6.6) نحصل على[13]:

$$\ln Y_t = \beta_1 + \beta_2 t + u_t \qquad (6.6.6)$$

هذا النموذج مثل أي نموذج انحدار خطي، حيث إن المعامل β_1 وβ_2 خطية. الفرق الوحيد هو أن المتغير المنحدر عليه هو لوغاريتم Y، والمتغير المنحدر هو «الزمن» والذي سيأخذ القيم 1، 2، 3 وهكذا.

نماذج مثل (6.6.6) تسمى نماذج شبه لوغاريتمية، حيث إن هناك متغيراً واحداً فقط (في هذه الحالة هذا المتغير هو المتغير المنحدر عليه) يظهر في صورة لوغاريتم. لتسهيل الوصف يسمى النموذج الذي يشتمل على المتغير المنحدر عليه في صورة لوغاريتم باسم نموذج Log–Lin. لاحقًا سنتحدث عن نموذج آخر يكون المتغير المنحدر عليه خطيًا والمتغيرات المنحدرة في صورة لوغاريتم، ويسمى هذا النموذج بنموذج Lin–Log.

قبل استعراض نتائج الانحدار، دعنا نناقش خصائص النموذج (5.6.6). في هذا النموذج يقيس معامل الميل النسبة الثابتة للتغير النسبي في Y بالنسبة للتغير المطلق المعطى في قيمة المتغير المنحدر (في هذه الحالة يعتبر هذا المتغير t). أي أن[14]

$$\beta_2 = \frac{\text{التغير النسبي في المتغير المنحدر عليه}}{\text{التغير المطلق في المتغير المنحدر}} \qquad (7.6.6)$$

(13) أضفنا حد الخطأ، لأن معادلة الفائدة المركبة لن تتحقق بالضبط. وأسباب إضافة حد الخطأ إلى التحويلة اللوغاريتمية مشروحة بالتفصيل في الفقرة 8.6.

(14) باستخدام التفاضل يمكن إثبات أن: $\beta_2 = d(\ln Y)/dX = (1/Y)(dY/dX) = (dY/Y)/dX$ والذي يعادل (7.6.6). إذا حدث تغير بسيط في Y و X فإن هذه العلاقة يمكن تقريبها لتصبح كالتالي:

$$\frac{(Y_t - Y_{t-1})/Y_{t-1}}{(X_t - X_{t-1})}$$

لاحظ أن $t = X$.

إذا ضربنا التغير النسبي في Y بـ 100، ستعطي (7.6.6) التغير النسبي، أو معدل النمو في Y بالنسبة للتغير المطلق في X (المتغير المنحدر) أي أن 100 مضروبة في β_2 تعطي معدل النمو في Y، 100 مضروبة في β_2 معروف تحت اسم شبه المرونة لـ Y بالنسبة لـ X (سؤال: حتى نحصل على المرونة، ما الذي يجب عمله؟).

مثال توضيحي: معدل نمو الإنفاق على الخدمات

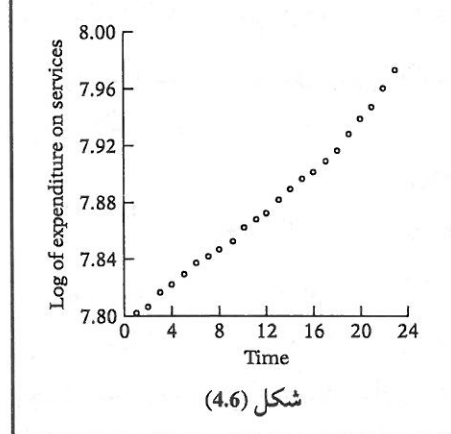
An Illustrative Example: The Rate of Growth Expenditure of Services

لتوضيح نموذج النمو (6.6.6). دعنا نعتبر بيانات الإنفاق على الخدمات المعطاة في جدول (3.6). نتائج الانحدار جاءت كالتالي:

$$\widehat{\ln EXS_t} = 7.7890 + 0.00743t$$
$$se = (0.0023) \quad (0.00017) \qquad (6.6.8)$$
$$t = (3387.619)^* \quad (44.2826)^* \quad r^2 = 0.9894$$

لاحظ أن: EXS ترمز إلى الإنفاق على الخدمات و* تعني أن قيمة P-value صغيرة للغاية.

لتفسير المعادلة (6.6.6) نرى أنه خلال الفترة الربع سنوية 1993:1 إلى 1998:3، الإنفاق على الخدمات زاد بمعدل (ربع سنوي) يساوي 0.743% وبالتالي ذلك تقريبًا يساوي معدل نمو سنوي 2.97%. بما أن 7.7890= لوغاريتم EXS عند بداية الدراسة، فإنه يأخذ معكوس اللوغاريتم نجد أن 2413.90 (بليون دولار) تساوي EXS عند بداية الدراسة (أي القيمة عند نهاية الربع الأخير من 1992). خط الانحدار الذي حصلنا عليه في المعادلة (8.6.6) موضح في الشكل (4.6).



شكل (4.6)

معدل النمو اللحظي في مقابل معدل النمو المركب:

Instantaneous versus Compound Rate of Growth

معامل متغير الاتجاه العام الموجود في نموذج النمو (6.6.6)، β_2 يمثل معدل النمو اللحظي (عند نقطة ما محددة) وليس معدل النمو المركب (خلال فترة زمنية). ولكن الأخير يمكن إيجاده بسهولة من (4.6.6) باستخدام معكوس اللوغاريتم للقيمة المقدرة لـ β_2، وطرح 1 منها ثم ضرب الفرق في 100. وبالتالي في مثالنا التوضيحي، القيمة المقدرة لمعامل الميل هي 0.00743 وبالتالي: [antilog (0.00743) − 1] = 0.00746 أو 0.746%

وبالتالي، في مثالنا التوضيحي، معدل النمو المركب للإنفاق على الخدمات يساوي تقريبًا 0.746% في الربع سنة، وهو أعلى قليلاً من معدل النمو اللحظي والمساوي لـ 0.743%. وهذا بالطبع يرجع إلى التأثير المركب.

نموذج الاتجاه العام الخطي : Linear Trend Model

بدلاً من تقدير النموذج (6.6.6)، يقوم الباحثون أحيانًا بتقدير النموذج التالي:

$$Y_t = \beta_1 + \beta_2 t + u_t \qquad (9.6.6)$$

أي أنه بدلاً من عمل انحدار للوغاريتم Y على الزمن، يقومون بعمل انحدار لـ Y على الزمن حيث Y يمثل المتغير المنحدر عليه محل الدراسة. مثل هذا النموذج يسمى نموذج الاتجاه العام الخطي، ويُعرف متغير الزمن t باسم متغير الاتجاه العام. إذا كان معامل الميل في (9.6.6) موجبًا فإن ذلك يعني وجود اتجاه عام متزايد في Y، أما إذا كان سالبًا فإن ذلك يعني وجود اتجاه عام متناقص.

بالنسبة لبيانات الإنفاق على الخدمات والتي سبق واستخدمتها، نتائج انحدار نموذج الاتجاه العام الخطي (9.6.9) جاءت كالتالي:

$$\begin{aligned} \widehat{\text{EXS}}_t &= 2405.848 \quad + \quad 19.6920t \\ t &= (322.9855) \quad (36.2479) \qquad r^2 = 0.9843 \end{aligned} \qquad (10.6.6)$$

على خلاف المعادلة (8.6.6) فإن تفسير (10.6.6) يكون كالتالي: خلال الفترة الربع سنوية 1993-I إلى 1998-III، فإنه في المتوسط يزداد الإنفاق على الاستهلاك بمعدل مطلق (وليس نسبيًا) يساوي 20 بليون دولار في الربع سنة. وبالتالي هناك اتجاه عام متزايد في الإنفاق على الخدمات.

الاختيار ما بين نموذج معدل النمو (8.6.6) ونموذج الاتجاه العام الخطي (10.6.6) يتوقف على ما إذا كان الباحث مهتمًا بالتغير النسبي أم التغير المطلق في نفقات الخدمات. بالطبع للمقارنة يكون من الأفضل استخدام التغير النسبي وليس المطلق. عمومًا لاحظ أننا لا نستطيع مقارنة قيمة r^2 للنموذج (8.6.6) مع نظيرها للنموذج (10.6.6) حيث إن المتغير المنحدر عليه مختلف في النموذجين.

سنرى في الفصل 7، كيف يمكن مقارنة R^2's للنماذج المختلفة مثل (8.6.6) و (10.6.6).

نموذج Lin–Log

على خلاف نموذج النمو السابق مناقشته، والذي نهتم فيه بدراسة النمو النسبي في Y بالنسبة لتغير مطلق في X، افترض أننا الآن مهتمون بالتغير المطلق في Y بالنسبة لتغير نسبي في X. النموذج الذي يمكننا من خلاله دراسة ذلك يمكن كتابته كالتالي:

$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i \qquad (11.6.6)$$

لتسهيل الوصف نسمي هذا النموذج بنموذج Lin-Log دعنا الآن نفسر معامل الميل β_2 [15].

$$\beta_2 = \frac{\text{التغير في } Y}{\text{التغير في } \ln X}$$

$$= \frac{\text{التغير في } Y}{\text{التغير النسبي في } X}$$

الخطوة الثانية تمت وفقًا لتساوي التغير في لوغاريتم أي قيمة بالتغير النسبي.

رمزيًا لدينا التالي:

$$\beta_2 = \frac{\Delta Y}{\Delta X/X} \qquad (12.6.6)$$

حيث، كالمعتاد Δ تعني تغيراً ضئيلاً. المعادلة (12.6.6) يمكن كتابتها بصورة مكافئة كالتالي:

$$\Delta Y = \beta_2(\Delta X/X) \qquad (13.6.6)$$

من هذه المعادلة، نجد أن التغير المطلق في Y(ΔY=) يساوي الميل مضروبًا في التغير النسبي في X. إذا كان الأخير مضروبًا في 100 فإن (13.6.6) تمثل التغير المطلق في Y بالنسبة للتغير النسبي لـ X. وبالتالي إذا تغيرت ($\Delta X/X$) بحوالي 0.01 وحدة (أو 1%) فإن التغير المطلق في Y سيساوي β_2(0.01)، وبالتالي عند التطبيق إذا وجدنا أن $\beta_2 = 500$، فإن التغير المطلق في Y سيساوي 0.5 = (500) (0.01). وبالتالي عند تقدير (11.6.6) باستخدام OLS، لا تنس ضرب القيمة المقدرة لمعامل الميل بـ0.01 أو أي قيمة أخرى لنفس المعلمة، ثم اقسم حاصل الضرب على 100. إذا لم تضع ذلك في الاعتبار، سيكون تفسيرك عند أي تطبيق غير سليم، ويعطي نتائج غير صحيحة.

---

(14) مرة ثانية، باستخدام التفاضل، لدينا

$$\frac{dY}{dX} = \beta_2\left(\frac{1}{X}\right)$$

وبالتالي

$$\beta_2 = \frac{dY}{\frac{dX}{X}} = (6.6.12)$$

السؤال العملي الآن: متى يكون استخدام نموذج lig–log كالموجود في (11.6.6) مفيد؟ تطبيق مهم، الإجابة على هذا السؤال متعلقة بما يسمى نماذج الإنفاق لـ Engel. وهذه النماذج سميت بذلك وفقًا للإحصائي الألماني Ernst Engel، 1821 إلى 1896 (انظر تمرين 10.6). افترض Engel التالي «النفقات الإجمالية للغذاء تتجه للزيادة بشكل رياضي، أما النفقات الإجمالية عمومًا فتزداد بشكل هندسي»[16].

مثال توضيحي AN ILLUSTRATIVE EXAMPLE

لتوضيح نموذج Lig–Log، دعنا نسترجع مثالنا الخاص بنفقات الغذاء في الهند، مثال (2.3). في هذا المثال تم توفيق نموذج خطي في المتغيرات كخطوة أولية. ولكن إذا قمنا برسم البيانات سنحصل على شكل (5.6).

من هذا الشكل، نجد أن نفقات الغذاء تزداد بدرجة أبطأ من زيادة النفقات الإجمالية، مما يؤيد مقولة Engel. نتائج نموذج Lin-Log وفقًا لهذه البيانات جاءت كالتالي:

$$\widehat{FoodExp}_i = -1283.912 + 257.2700 \ln TotalExp_i$$
$$t = (-4.3848)^* \quad (5.6625)^* \quad r^2 = 0.3769 \qquad (14.6.6)$$

لاحظ أن: * تعني أن قيمة P-value صغيرة للغاية.

Expenditure on food (Rs.) — Total expenditure (Rs.)

شكل (5.6)

ولتفسير النتائج كالمعتاد، نجد أن معامل الميل، والمساوي لـ 257 تقريبًا يعني أن لكل زيادة في إجمالي النفقات بـ 1% فإنه في المتوسط يزداد إجمالي نفقات الغذاء بـ 2.57 وفقًا للـ 55 أسرة الموجودة في العينة.

(لاحظ أننا: قسمنا معامل الميل المقدر على 100).

قبل الانتقال إلى نقطة أخرى، لاحظ أنه إذا أردت الحصول على معامل المرونة لنماذج Log-Lin أو Lin-Log، يمكنك استخدام تعريف معامل المرونة والمعرف كالتالي:

$$\text{المرونة} = \frac{dY}{dX}\frac{X}{Y}$$

في واقع الأمر، بمجرد معرفة شكل الدالة الخاصة بالنموذج، يمكن بسهولة حساب المرونة بتطبيق التعريف السابق.

(جدول (6.6) المعطى سابقًا معطى فيه تلخيص لمعاملات المرونة الخاصة بالنماذج المختلفة).

---

(16) انظر Chandan Mukherjee, Howard White, and Marec Wuyts, Econometrics and data Analysis for Developing Countries, Routledge, London, 1988, p. 158. This quote is attributed to H. Working, "Statistical Laws of Family Expenditure, "Journal of the American Statistical Association, vol. 38, 1943, pp. 43-56.

7.6 نماذج المقلوب : RECIPROCAL MODELS

النماذج التي لها الشكل التالي تعرف باسم نماذج المقلوب .

$$Y_i = \beta_1 + \beta_2 \left(\frac{1}{X_i} \right) + u_i \qquad (1.7.6)$$

على الرغم من أن هذا النموذج غير خطي في المتغير X، حيث إنه موجود في صورة المقلوب أو المعكوس، إلا أن هذه النماذج خطية في β_1 و β_2 وبالتالي تعتبر نماذج انحدار خطية[17] .

هذا النموذج لديه الخصائص التالية : كلما زادت X فإن المقدار $\beta_2(1/X)$ يؤول إلى الصفر (لاحظ أن : β_2 ثابت) وتؤول Y تقاريبًا إلى β_1 . وبالتالي نماذج مثل (1.7.6) يوجد فيها قيمة تقاربية أو نهاية خاصة بالمتغير التابع سيصل إليها إذا زاد المتغير X بشكل لا نهائي[18] .

بعض المنحنيات المحتملة لـ (1.7.6) موضحة في شكل (6.6) . كتوضيح لشكل (a6.6) اعتبر البيانات المعطاة في جدول (4.6) . هذه البيانات المقطعية خاصة بـ 64 دولة، وهي بيانات خاصة بوفيات الأطفال وبعض المتغيرات الأخرى . الآن اهتم فقط بالمتغيرات التالية :

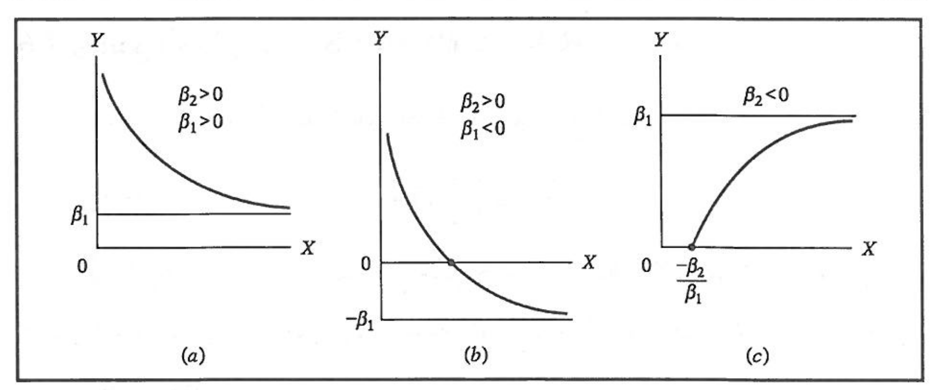
وفيات الأطفال (CM) وGNP والمرسومات في الشكل (7.6) .

كما ترى في شكل (a6.6) : كلما زاد GNP للفرد، فإننا نتوقع انخفاض وفيات الأطفال، حيث إن الأشخاص ستكون لديهم قدرة مالية أعلى للإنفاق على الاهتمام بالصحة، وذلك طبعًا بافتراض ثبات باقي العوامل الأخرى المؤثرة . ولكن هذه العلاقة لا تأخذ شكل الخط المستقيم : فكلما زاد GNP فإنه يحدث انخفاض ملحوظ مبدئيًا في CM، ثم يقل هذا الانخفاض كلما زاد GNP الخاص بالدول المختلفة .

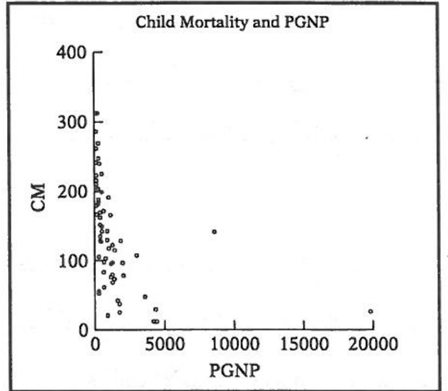
---

(17) إذا استخدمنا $X_i^* = (1/X_i)$ ، فإن (1.7.6) سيكون خطيًا في المعالم وخطيًا أيضًا في المتغيرات Y_i و X_i^* .

(18) الميل في (1.7.6) يساوي : $dY/dX = -\beta_2(1/X^2)$ ، مما يعني انه إذا كان β_2 موجبًا فإن الميل سالب وإذا كان β_2 سالبًا سيكون الميل موجبًا . انظر الأشكال (a6.6) و (c6.6) بالترتيب .



شكل (6.6) النموذج المقلوب : $Y = \beta_1 + \beta_2 (\frac{1}{X})$



شكل (7.6) العلاقة بين وفيات الأطفال وGNP بالنسبة للفرد في 66 دولة

اذا أردت توفيق نموذج المقلوب (1.7.6) للبيانات السابقة، ستكون نتائج الانحدار كالتالي :

$$\widehat{\text{CM}}_i = 81.79436 + 27{,}273.17\left(\frac{1}{\text{PGNP}_i}\right)$$

$$\text{se} = (10.8321) \quad (3759.999) \qquad \textbf{(2.7.6)}$$

$$t = (7.5511) \quad (7.2535) \quad r^2 = 0.4590$$

كلما زاد GNP للفرد ووصل إلى قيمة لا نهائية، فإن وفيات الأطفال ستؤول تقاريبًا إلى قيمة تساوي 82 وفاة لكل ألف. كما وضحنا في هامش 18، القيمة الموجبة لمعامل (1/PGNA) يعني أن معدل التغير في CM بالنسبة لـ PGNP سالب.

أحد التطبيقات المهمة لشكل (b6.6) هو منحنيات الاقتصاد الكلي لـ Phillips. باستخدام البيانات الخاصة بمعدل تغير الأجور (Y) ومعدل البطالة (X) في المملكة

المتحدة خلال الفترة 1861 إلى 1957. حصل Phillips على منحنى شكله العام يتبع شكل (b6.6) (شكل 8.6)[19].

كما يتضح من شكل (8.6)، هناك تماثل في تغير الأجور وفقًا لمستوى معدل البطالة: الأجور تزداد أسرع وفقًا للتغير في معدل البطالة إذا كان معدل البطالة أقل من U_n، والذي يسمه الاقتصاديون بالمعدل الطبيعي للبطالة [ويعرف على أنه معدل البطالة المطلوب للحفاظ على التضخم (الأجور) ثابت].

جدول (4.6) الخصوبة وبيانات أخرى خاصة بـ 64 دولة

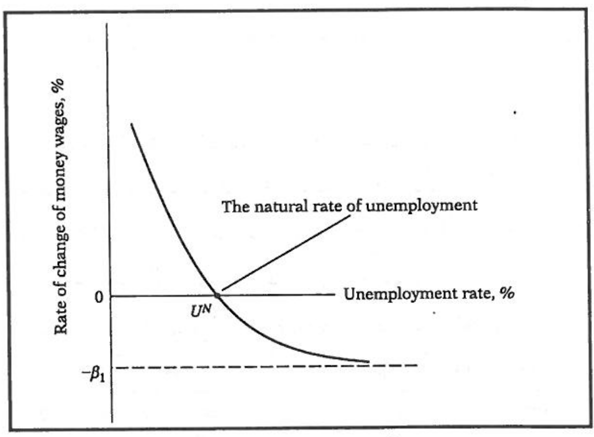
Observation	CM	FLFP	PGNP	TFR	Observation	CM	FLFP	PGNP	TFR
1	128	37	1870	6.66	33	142	50	8640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1620	3.91
5	96	76	2050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1180	3.93	43	191	31	1010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1730	3.50	45	37	88	1730	3.46
14	165	31	1150	7.41	46	103	35	780	5.66
15	94	77	1160	4.21	47	67	85	1300	4.82
16	96	80	1270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1340	7.17
24	12	81	4240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4370	2.86
27	107	87	3020	6.66	59	121	41	1310	4.88
28	72	63	1420	7.28	60	115	62	1470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19830	5.23	62	47	85	3630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

لاحظ أن CM = وفيات الأطفال، عدد وفيات الأطفال أقل من 5 سنوات في السنة لكل 1000 مولد حي.
FLFP = معدل الأمية في النساء.
PGNP = GNP مفرد في 1980.
TFR = معدل الخصوبة الكلي، 1980–1985، متوسط عدد الأطفال الذي تنجبه الأنثى وفقًا لمعدل خصوبة معين في سنة ما.

المصدر: Chandan Mukherjee, Howard White, and Marc Whyte, Econometrics and Data Analysis for Developing Countries, Routledge, London, 1998, p. 456.

---

(19) A. W. Phillips, "The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957," Economica, November 1958, vol. 15, pp. 283-299.
لاحظ أن المنحنى الأصلي لا يتقاطع مع محور معدل البطالة، لكن شكل (8.6) يمثل مرحلة تالية للمنحنى.



شكل (8.6) منحنى Phillips

وتنخفض بمعدل متساو عندما يكون معدل البطالة أعلى من المعدل الطبيعي، β_1، مما يعني أن القيمة التقاربية للأجور تغيرت.

هذه الخاصية المتعلقة بمنحنى philips قد ترجع إلى عوامل أخرى مؤسسية، مثل قوة الشراء، الحد الأدنى للأجور، تعويضات البطالة وهكذا. ونظرًا لأنه منذ مقالة philips وجد العديد من الدراسات التي تناولت منحنى phillips سواء من الجانب النظري أو التطبيقي. فلا يوجد متسع من حيث الوقت أو المكان لاستعراض كافة التفاصيل المتعلقة لهذا الموضوع في إطار الكتاب الحالي. فمنحنى phillips مر بالعديد من المراحل، وكصورة أخيرة معدلة انظر في Olivier Blanchard.[20] إذا افترضنا أن π_t ترمز إلى معدل التضخم عند الزمن t وعرفناه على أنه معدل التغير في مستوى السعر وفقًا لمؤشر سعر ما، مثل مؤشر سعر المستهلك (CPI) وUN_t ترمز إلى معدل البطالة عند الزمن t، فإن الشكل التالي يمثل صورة حديثة لمنحنى phillips كالتالي:

$$\pi_t - \pi_t^e = \beta_2(UN_t - U^n) + u_t \qquad (3.7.6)$$

حيث π_t = معدل التضخم الفعلي عند الزمن t.

π_t^e = معدل التضخم المتوقع عند الزمن t، التوقع تم في الزمن $(t-1)$.

UN_t = معدل البطالة الفعلي عند الزمن t.

U^n = معدل البطالة الطبيعي عند الزمن t.

u_t = حد خطأ عشوائي[21].

(20) انظر Olivier Blanchard, Macroeconomics, Prentice Hall, Englewood Cliffs, N.J., 1997, Chap. 17

(21) يعتبر الاقتصاديون أن حد الخطأ هذا يمكن اعتباره نوعًا من تقلبات العرض أو صدماته، مثل أزمة البترول وفقًا للـ OPEC خلال الفترة من 1973 إلى 1979.

بما أن π_t^e لا يمكن ملاحظته مباشرةً، فكنقطة مبدئية يمكن الاعتماد على الفرض المبسط القائل بأن $\pi_t^e = \pi_{t-1}$، أي أن التضخم المتوقع هذا العام هو معدل التضخم الموجود في العام السابق. بالطبع يمكن وجود فروض أكثر تعقيداً من ذلك، وسنناقش هذا الموضوع بالتفصيل في الفصل 17.

بالتعويض عن هذا الافتراض في (3.7.6)، وكتابة نموذج الانحدار في صورته التقليدية، نحصل على المعادلة المقدرة التالية:

$$\pi_t - \pi_{t-1} = \beta_1 + \beta_2 \text{UN}_t + u_t \qquad \textbf{(4.7.6)}$$

حيث $\beta_1 = \beta_2 U^n$. المعادلة (4.7.6) تعني أن التغير في معدل التضخم بين فترتين زمنيتين مرتبط خطيًا بمعدل البطالة الحالي. مبدئيًا β_2 يتوقع أن تكون سالبة (لماذا؟) و β_1 يتوقع أن تكون موجبة (وذلك وفقًا لأن β_2 سالبة و U_n موجبة).

ويظهر هنا أن علاقة Phillips المعطاة في (3.7.6) والتي تعرف باسم منحنى Phillips المعدل أو منحنى Phillips للتوقعات المساعدة (وذلك لتوضيح أن π_{t-1} يعني التضخم المتوقع) أو منحنى Phillips المعجل (وذلك حيث إن معدل البطالة المنخفض يؤدي إلى زيادة في معدل التضخم، وبالتالي تزداد مستويات الأسعار).

لتوضح منحنى Phillips المعدل، دعنا نستخدم بيانات جدول (5.6) والخاصة بالتضخم مقاس سنة بسنة كنسبة من مؤشر سعر المستهلك (CPIflation) ومعدل البطالة خلال الفترة 1960 – 1998.

معدل البطالة يمثل معدل البطالة المدينة. من هذه البيانات نحصل على التغير في معدل التضخم $(\pi_t - \pi_{t-1})$ ثم رسمه ضد معدل البطالة، ونحن هنا نستخدم CPI كمقياس للتضخم. الشكل الناتج موجود في شكل (9.6).

كالمتوقع العلاقة بين التغير في معدل التضخم ومعدل البطالة، تعتبر علاقة عكسية – فمعدل البطالة المنخفض يؤدي إلى زيادة في معدل التضخم، وبالتالي تزداد مستويات الأسعار، وهنا سمي المنحنى بمنحنى Phillip المعجل.

بالنظر إلى شكل (9.6)، لا يتضح ما إذا كان من الأفضل استخدام نموذج انحدار خطي (خط مستقيم) أو نموذج انحدار المقلوب، حيث إن العلاقة الخطية والمنحنية محتملة بين المتغيرين.

وبالتالي نستعرض الآن نتائج الانحدار بناء على النموذجين معًا عمومًا، ضع في الاعتبار أن في نموذج انحدار المقلوب نتوقع أن يكون الجزء الثابت المقطوع من المحور الصادي سالبًا والميل موجبًا، كما سبق وأوضحنا ذلك في هامش (18).

النموذج الخطي:

$$\widehat{(\pi_t - \pi_{t-1})} = 4.1781 - 0.6895\ \text{UN}_t$$
$$t = (3.9521) \quad (-4.0692) \qquad r^2 = 0.3150 \qquad (5.7.6)$$

جدول (5.6) معدل التضخم ومعدل البطالة، الولايات المتحدة، 1960 – 1998

Observation	INFLRATE	UNRATE	Observation	INFLRATE	UNRATE
1960	1.7	5.5	1980	13.5	7.1
1961	1.0	6.7	1981	10.3	7.6
1962	1.0	5.5	1982	6.2	9.7
1963	1.3	5.7	1983	3.2	9.6
1964	1.3	5.2	1984	4.3	7.5
1965	1.6	4.5	1985	3.6	7.2
1966	2.9	3.8	1986	1.9	7.0
1967	3.1	3.8	1987	3.6	6.2
1968	4.2	3.6	1988	4.1	5.5
1969	5.5	3.5	1989	4.8	5.3
1970	5.7	4.9	1990	5.4	5.6
1971	4.4	5.9	1991	4.2	6.8
1972	3.2	5.6	1992	3.0	7.5
1973	6.2	4.9	1993	3.0	6.9
1974	11.0	5.6	1994	2.6	6.1
1975	9.1	8.5	1995	2.8	5.6
1976	5.8	7.7	1996	3.0	5.4
1977	6.5	7.1	1997	2.3	4.9
1978	7.6	6.1	1998	1.6	4.5
1979	11.3	5.8			

لاحظ أن: معدل التضخم هو التغير النسبي سنة إلى سنة في CPI، معدل البطالة لمثل معدل بطالة المدينين.

المصدر: Economic Report of the President, 1999, Table B-63, p. 399, for CPI changes and Table B-42, p. 376, for the unemployment rate.



شكل (9.6) منحنى Phillips المعدل

النموذج المقلوب :

$$\widehat{(\pi_t - \pi_{t-1})} = -3.2514 + 18.5508\left(\frac{1}{\mathrm{UN}_t}\right) \qquad \textbf{(6.7.6)}$$
$$t = (-2.9715) \quad (3.0625) \quad r^2 = 0.2067$$

كل المعاملات المقدرة في كل من النموذجين معنويين إحصائيًا، حيث إن كل قيم الـ P-values أقل من 0.05 .

نموذج (5.7.6) يوضح أنه إذا انخفض معدل البطالة بـ 1% فإنه في المتوسط التغير في معدل التضخم يزداد بحوالي 0.7% والعكس صحيح .

نموذج (6.7.6) يوضح أنه إذا زاد معدل البطالة بشكل غير محدود، فإن التضخم سينخفض بحوالي 3.25% . وبالتالي من المعادلة (5.7.6) يمكن حساب معدل البطالة الطبيعي كالتالي :

$$U^n = \frac{\hat{\beta}_1}{-\hat{\beta}_2} = \frac{4.1781}{0.6895} = 6.0596 \qquad \textbf{(7.7.6)}$$

أي أن معدل البطالة الطبيعي يساوي تقريبًا 6.06% . وبالتالي يضع الاقتصاديون المعدل الطبيعي بين 5% إلى 6%، على الرغم من أنه في السنوات السابقة في الولايات المتحدة كان المعدل الطبيعي أقل من ذلك بكثير .

النموذج اللوغاريتمي المقلوب :

Log Hyperbola or Logarithmic Reciprocal Model

لاستنتاج خلاصة مناقشتنا السابقة للنماذج المقلوبة . دعنا نعتبر نموذج الانحدار التالي، والمسمى بالنموذج اللوغاريتمي المقلوب والذي يمكن كتابته كالتالي :

$$\ln Y_i = \beta_1 - \beta_2\left(\frac{1}{X_i}\right) + u_i \qquad \textbf{(8.7.6)}$$

الشكل البياني لهذا النموذج معروض في شكل (10.6) . كما يتضح من هذا الشكل، فإن Y تزيد بمعدل متزايد (حيث إن المنحنى مقعر) ثم يزداد بمعدل متناقص (حيث إن المنحنى يصبح محدبًا)[22] .

---

(22) وفقاً لقواعد الحساب يمكن إثبات التالي :

$$\frac{d}{dX}(\ln Y) = -\beta_2\left(-\frac{1}{X^2}\right) = \beta_2\left(\frac{1}{X^2}\right)$$

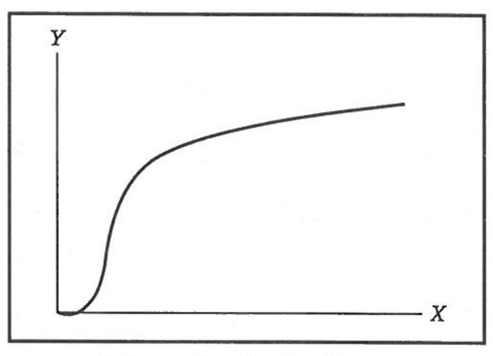
ولكن

$$\frac{d}{dX}(\ln Y) = \frac{1}{Y}\frac{dY}{dX}$$

بالتعويض نحصل على

$$\frac{dY}{dX} = \beta_2\frac{Y}{X^2}$$

والذي يمثل ميل Y بالنسبة لـ X .



شكل (10.6) النموذج اللوغاريتمي المقلوب

وبالتالي، قد يكون النموذج مناسبًا لتمثيل دالة إنتاج قصيرة الأجل. تذكر أنه وفقًا للاقتصاد الجزيئي إذا كان رأس المال والعمالة من مدخلات دالة الإنتاج، وإذا اعتبرنا رأس المال ثابتًا وزاد مدخل العمالة، فإن العلاقة الممثلة للعمالة والناتج في الفترة قصيرة الأجل ستتطابق مع شكل (10.6) (انظر مثال (3.7)، الفصل 7).

8.6 اختيار شكل الدالة : CHOICE OF FUNCTIONAL FORM

في هذا الفصل، ناقشنا العديد من أشكال الدوال المختلفة وفقًا لنماذج تطبيقية تم افتراضها داخل إطار نماذج لانحدار خطية المعالم. اختيار شكل دالة معينة قد يكون سهلاً في حالة وجود متغيرين اثنين فقط، حيث إننا يمكن أن نرسم المتغيرين، وبالتالي نكون فكرة مبدئية عن النموذج المناسب. الاختيار يصبح أكثر صعوبة عندما يكون لدينا نموذج انحدار متعدد يشتمل على أكثر من متغير منحدر واحد، وسنرى ذلك بالتفصيل عند مناقشة هذا الموضوع في الفصلين القادمين. ولا يمكن إنكار أن عملية اختيار النموذج المناسب للتقدير التطبيقي تحتاج إلى مهارة خاصة وخبرة جيدة.

ولكن يمكننا هنا إعطاء بعض الإرشادات الممكن اتباعها :

1 - النظرية الموجودة وراء البحث محل الدراسة (مثل منحنى Phillips) قد تقترح شكل دالة معين.

2 - من المفيد إيجاد معدل التغير (أي الميل) الخاص بالمتغير المنحدر عليه بالنسبة للمتغير المنحدر، وإيجاد المرونة أيضًا. بالنسبة للنماذج المختلفة التي تم استعراضها في هذا الفصل، سنقدم في جدول (6.6) أشكال الدوال المناسبة لمعاملات الميل والمرونة. معرفة هذه الأشكال قد تفيدك عند التطبيق على نماذج أخرى.

جدول (6.6)

Model	Equation	Slope $\left(=\frac{dY}{dX}\right)$	Elasticity $\left(=\frac{dY}{dX}\frac{X}{Y}\right)$
Linear	$Y=\beta_1+\beta_2X$	β_2	$\beta_2\left(\frac{X}{Y}\right)^*$
Log–linear	$\ln Y=\beta_1+\beta_2\ln X$	$\beta_2\left(\frac{Y}{X}\right)$	β_2
Log–lin	$\ln Y=\beta_1+\beta_2 X$	$\beta_2\,(Y)$	$\beta_2\,(X)^*$
Lin–log	$Y=\beta_1+\beta_2\ln X$	$\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{1}{Y}\right)^*$
Reciprocal	$Y=\beta_1+\beta_2\left(\frac{1}{X}\right)$	$-\beta_2\left(\frac{1}{X^2}\right)$	$-\beta_2\left(\frac{1}{XY}\right)^*$
Log reciprocal	$\ln Y=\beta_1-\beta_2\left(\frac{1}{X}\right)$	$\beta_2\left(\frac{Y}{X^2}\right)$	$\beta_2\left(\frac{1}{X}\right)^*$

3 - معاملات النموذج المختارة، يجب أن توضع وفقًا لتوقعات مسبقة. فمثلاً، إذا كنا بصدد دراسة الطلب على السيارات، وافترضنا علاقة بينه وبين السعر وبعض المتغيرات الأخرى، علينا أن نتوقع أن يكون معامل متغير السعر سالبًا.

4 - أحيانًا يوجد أكثر من نموذج واحد مناسب للبيانات. في منحنى Phillips المعدل، قمنا باستخدام النموذج الخطي، والنموذج المقلوب لنفس البيانات. وفي كلتا الحالتين المعاملات كانت منطقية ومتماشية مع التوقعات المسبقة وكانتا معنوتتين إحصائيًا. فرق واحد مهم خاص بقيمة r^2 والتي كانت أكبر في حالة النموذج الخطي عن نموذج المقلوب مما يعني أفضلية لاستخدام النموذج الخطي عن نموذج المقلوب للبيانات محل الدراسة. ولكن ضع دائمًا في الاعتبار عند مقارنة قيم r^2 الاثنين أن يكون المتغير التابع (المنحدر عليه) هو نفس المتغير في النموذجين، أما المتغيرات المنحدرة فممكن أن تختلف تمامًا. سنلقى مزيدًا من الضوء على تلك النقطة في الفصل القادم.

5 - عمومًا يجب على الباحث ألا يزيد من أهمية r^2، بمعنى أن r^2 الأعلى تعني بالضرورة نموذج أفضل. كما سنرى في الفصل القادم، فإن r^2 تزيد كلما أضفنا مزيدًا من المتغيرات المنحدرة إلى النموذج. فالأهم هو النظرية المبنى عليها النموذج المختار، وأشارت المعاملات قدرة ومعنويتهم الإحصائية. فإذا كان النموذج جيدًا وفقًا لما سبق، حتى وإن كانت r^2 منخفضة، فإن النموذج يكون مناسبًا للبيانات. سنعود مرة أخرى إلى هذا الموضوع المهم بمزيد من التفاصيل في الفصل 13.

9.6* ملاحظة خاصة بطبيعة حد الخطأ العشوائي:
A NOTE OF THE NATURE OF THE STOCHASTIC ERROR TERM:
حد الخطأ العشوائي التجميعي في مقابل الضربي : ADDITIVE VERSUS MULTIPLICATIVE STOCHASTIC ERROR TERM

اعتبر نموذج الانحدار التالي والمماثل لـ (1.5.6) ولكن بدون حد الخطأ:

$$Y_i = \beta_1 X^{\beta_2} \quad (1.9.6)$$

لتسهيل التقدير، دعنا نعبر عن هذا النموذج بالثلاثة أشكال المختلفة التالية:

$$Y_i = \beta_1 X_i^{\beta_2} u_i \quad (2.9.6)$$

$$Y_i = \beta_1 X_i^{\beta_2} e^{u_i} \quad (3.9.6)$$

$$Y_i = \beta_1 X_i^{\beta_2} + u_i \quad (4.9.6)$$

ادخل اللوغاريتم على طرفي المعادلات السابقة تحصل على:

$$\ln Y_i = \alpha + \beta_2 \ln X_i + \ln\ u_i \quad (2a.9.6)$$

$$\ln Y_i = \alpha + \beta_2 \ln X_i + u_i \quad (3a.9.6)$$

$$\ln Y_i = \ln\left(\beta_1 X_i^{\beta_2} + u_i\right) \quad (4a.9.6)$$

حيث $\alpha = \ln \beta_1$

نماذج مثل (2.9.6) تعتبر نماذج انحدار خطية جوهريًا (في المعالم). بمعنى أنه باستخدام التحويلة المناسبة (اللوغاريتم) يصبح النموذج خطيًا في المعالم α و β_2. (لاحظ أن: هذه النماذج غير خطية في β_1) ولكن النموذج (4.9.6) هو غير خطي جوهريًا في المعالم. فلا توجد طريقة بسيطة لاستخدام اللوغاريتم لـ (4.9.6) حيث إن $\ln (A + \beta) \neq \ln A + \ln \beta$.

على الرغم من أن (2.9.6) و (3.9.6) نماذج انحدار خطية، ويمكن تقديرهما بسهولة باستخدام OLS أو ML، إلا أنه يجب التعامل بحذر مع خصائص حد الخطأ العشوائي الموجود في هذه النماذج. تذكر أن خاصة BLUE للـ OLS تتطلب أن تكون القيمة المتوقعة لـ u_i تساوي الصفر، ويكون التباين ثابتًا ولا يوجد ارتباط ذاتي.

(*) اختياري.

لاختيارات الفروض نضع مزيداً من الفروض حول u_i، فيجب أن يتبع التوزيع الطبيعي بتوقع يساوي الصفر وبتباين كالسابق ذكره. باختصار نفترض أن $u_i \sim N(0, \sigma^2)$.

الآن اعتبر النموذج (2.9.6). نظيره الإحصائي معطى في (2a.9.6) لاستخدام نموذج الانحدار الخطي الطبيعي التقليدي (CNLRM) يجب أن نفترض أن:

$$\ln u_i \sim N(0, \sigma^2) \qquad (5.9.6)$$

وبالتالي عندما نقوم بعمل انحدار (2a.9.6) يجب أن نطبق اختبارات الاعتيادية السابق مناقشتها في الفصل 5 على البواقي التي نحصل عليها من الانحدار. لاحظ أنه إذا كان $\ln u_i$ يتبع التوزيع الطبيعي بتوقع يساوي الصفر وتباين ثابت، فإن النظرية الإحصائية تثبت أن u_i الموجود في (2.9.6) يجب أن يتبع التوزيع الطبيعي اللوغاريتمي بتوقع يساوي $e^{\sigma^2/2}$ وتباين $e^{\sigma^2}(e^{\sigma^2}-1)$.

كما يتضح من التحليل السابق، يجب على الباحث أن يضع في الاعتبار حد الخطأ خصوصاً عند تحويل النموذج لتحليل الانحدار.

فمثلاً في (4.9.6) فإن هذا النموذج هو نموذج انحدار غير خطي في المعالم، ويجب حله من خلال بعض العمليات التكرارية باستخدام الحاسب الآلي. أما في نموذج (3.9.6) فلا يوجد فيه أي مشاكل مسبقة عند التقدير.

عموماً في النهاية ضع في اعتبارك حد الخطأ عندما تقوم بعمل تحويل للنموذج لتحليل الانحدار. أما التطبيق الأعمى لـ OLS على النموذج المحول سيؤدي إلى نموذج لا توجد فيه الخصائص الإحصائية المرغوب فيها.

10.6 الملخص والاستنتاجات : SUMMARY AND CONCLUSIONS

في هذا الفصل، قدمنا العديد من النقاط المهمة والمتعلقة بنموذج الانحدار الخطي التقليدي (CLRM).

1 - أحياناً لا يشتمل نموذج الانحدار على جزء ثابت مقطوع من المحور الصادي. مثل هذه النماذج تعرف باسم النماذج المارة بنقطة الأصل. على الرغم من أن الخطوات الجبرية الخاصة بتقدير هذه النماذج تعتبر خطوات بسيطة، إلا أنه يجب على الباحث أن يستخدم هذه النماذج بمزيد من الحذر. ففي هذه النماذج يكون مجموع البواقي $\sum\hat{u}_i$ لا يساوي الصفر، بالإضافة إلى أن قيمة r^2 المحسوبة قد لا

يكون لها أي معنى. وبالتالي إذا لم يكن هناك ضرورة نظرية لافتراض عدم وجود جزء ثابت، فإنه من الأفضل أن يوجد في نموذج الانحدار بشكل صريح الجزء الثابت المقطوع من المحور الصادي.

2 - الوحدات والموازين المستخدمة لقياس المتغير المنحدر عليه، والمتغيرات المنحدرة مهمة جدًا، حيث إن تفسير معاملات النموذج يعتمد بشكل كبير على هذه الوحدات. في الأبحاث التطبيقية يجب على الباحث ليس فقط توضيح مصدر بياناته، ولكن أيضًا يجب عليه أن يوضح الوحدات المستخدمة لقياس المتغيرات.

3 - شكل الدالة الممثلة للعلاقة بين المتغير المنحدر عليه والمتغيرات المنحدرة يعتبر من الأمور المهمة والتي لا تقل أهمية عن كل ما سبق.

بعض أشكال الدوال المهمة والسابق ذكرها في هذا الفصل هي (a) نموذج المرونة الثابت، أو النموذج الخطي اللوغاريتمي، (b) نماذج الانحدار شبه اللوغاريتمية و(c) نماذج المقلوب.

4 - في النماذج الخطية اللوغاريتمية يكون كلاً من المتغير المنحدر عليه، والمتغيرات المنحدرة في صورة لوغاريتمية. معاملات الانحدار عليه بالنسبة للمتغير المنحدر.

5 - في النماذج شبه اللوغاريتمية يكون إما المتغير المنحدر عليه أو المتغير المنحدر في صورة لوغاريتمية. في النماذج شبه اللوغاريتمية عندما يكون المتغير المنحدر عليه هو المتغير الموجود في صورة لوغاريتمية يكون المتغير المنحدر عليه هو الزمن، ومعامل الميل المقدر (مضروب في 100) يقيس معدل النمو (اللحظي) في المتغير المنحدر عليه.

مثل هذه النماذج عادة ما تستخدم لقياس معدلات النمو في العديد من الظواهر الاقتصادية. في النماذج شبه اللوغاريتمية عندما يكون المتغير المنحدر هو المتغير الموجود في صورة لوغاريتمية، فإن معامله يقيس معدل التغير المطلق في المتغير المنحدر عليه وفقًا للتغير النسبي في قيمة المتغير المنحدر.

6 - في نماذج المقلوب، أما المتغير المنحدر عليه أو المتغير المنحدر أي منهما يكون في صورة المقلوب أو المعكوس للتعبير عن العلاقة غير الخطية بين المتغيرات الاقتصادية كالموجودة في منحنى Phillips.

7 - لاختيار شكل الدالة، يجب أن نهتم جيدًا بحد الخطأ العشوائي u_i، فكما سبق وذكرنا في الفصل 5، فإن CLRM يفترض صراحةً أن حد الخطأ له توقع يساوي الصفر، وتباين ثابت (ثبات التباين) ويكون غير مرتبط بالمتغير المنحدر. وفقًا

لهذه الفروض فإن مقدرات OLS تعتبر BLUE. أيضًا وفقًا لـ CNLRM، فإن مقدرات OLS تتبع التوزيع الطبيعي. يجب على الباحث أن يتحقق من أن الفروض المطلوبة مستوفاة في شكل الدالة المختار للتحليل التطبيقي.

بعد أن يقوم الباحث بعمل الانحدار يجب أن يستخدم بعض الاختبارات للتحقق من النموذج، مثل اختبار الاعتيادية السابق مناقشته في الفصل 5. أما بالنسبة لاختبارات الفروض التقليدية مثل اختبارات t، F و χ^2 فلابد من التأكد من أن حد الخطأ يتبع التوزيع الطبيعي خصوصًا إذا كان حجم العينة صغيراً.

8 - على الرغم من أن مناقشتنا مقتصرة على نموذج الانحدار ثنائي المتغيرات، فإنه في الفصول القادمة سنرى أن الحالات الأخرى التي يوجد فيها أكثر من متغيرين داخل نموذج الانحدار ما هي إلا امتداد لحالة النموذج ثنائي المتغيرات، ولكنها تحتوي على مزيد من الخطوات الجبرية التي يصعب استعراضها عند التقويم الأولي لنماذج الانحدار. وهذا هو السبب الذي يجعل من الضروري استعراض نماذج الانحدار ثنائية المتغيرات.

تمارين : EXERCISES

أسئلة : Questions

1.6 اعتبر نموذج الانحدار التالي:

$$y_i = \beta_1 + \beta_2 x_i + u_i$$

حيث $y_i = Y_i - \bar{Y}$ و $x_i = X_i - \bar{X}$. في هذه الحالة، يمر خط الانحدار على نقطة الأصل. هل العبارة السابقة صح أم خطأ؟ وضح جميع الخطوات الحسابية المتبعة.

2.6 نتائج الانحدار التالية حصلنا عليها بناء على بيانات شهرية خلال الفترة يناير 1978 إلى ديسمبر 1987(*).

$$\begin{aligned} \hat{Y}_t &= 0.00681 \quad + 0.75815X_t \\ \text{se} &= (0.02596) \quad (0.27009) \\ t &= (0.26229) \quad (2.80700) \\ p \text{ value} &= (0.7984) \quad (0.0186) \qquad r^2 = 0.4406 \end{aligned}$$

(*) البيانات محل الدراسة تم الحصول عليها من بيانات diskette الموجود في Ernst R. Berndt, The Practice of Econometrics: Cllassic and Contemporary, Addison-Wesley, Reading, Mass., 1991.

$$\hat{Y}_t = 0.76214X_t$$
$$se = (0.265799)$$
$$t = (2.95408)$$
$$p \text{ value} = (0.0131) \qquad r^2 = 0.43684$$

حيث Y = معدل العائد الشهري على أسهم Texaco ، % و X = معدل عائد السوق الشهري %.

(a) ما الفرق بين الانحدارين السابقين؟

(b) بناء على النتائج السابقة ، هل من الأفضل الاحتفاظ بجزء ثابت مقطوع من المحور الصادي في النموذج الأول؟ علل إجابتك .

(c) كيف يمكنك تفسير معاملات الميل في النموذجين السابقين؟

(d) ما هي النظرية المستخدمة وراء كل من النموذجين السابقين؟

(e) هل يمكنك مقارنة r^2 الخاصة بالنموذجين السابقين؟ علل إجابتك .

(f) إحصاء الاعتيادية لـ Jarque-Bera للنموذج الأول في هذه المسئلة يمكن أن تستنتجه من هذه الإحصاءات؟

(g) قيمة t الخاصة بمعامل الميل في النموذج المار بنقطة الأصل تساوي تقريبًا 2.95 ، وفي النموذج الذي يشتمل على جزء ثابت تساوي تقريبًا 2.81 . هل يمكنك تفسير هذه النتيجة؟

3.6 اعتبر نموذج الانحدار التالي :

$$\frac{1}{Y_i} = \beta_1 + \beta_2 \left(\frac{1}{X_i}\right) + u_i$$

لاحظ أن كلاً من Y و X يفترض ألا يتساويان مع الصفر .

(a) هل هذا النموذج يعتبر نموذج انحدار خطي؟

(b) كيف يمكنك تقدير هذا النموذج؟

(c) ما الذي سيحدث للـ Y عندما تؤول X إلى ما لا نهاية؟

(d) هل يمكنك إعطاء مثال يكون النموذج السابق نموذجًا مناسبًا له؟

4.6 اعتبر النموذج الخطي اللوغاريتمي التالي :

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

ارسم Y على المحور الرأسي و X على المحور الأفقي. ارسم المنحنى الموضح للعلاقة بين Y و X عندما يكون $\beta_2 = 1$ وعندما يكون $\beta_2 > 1$ وعندما يكون $\beta_2 < 1$.

5.6 اعتبر النماذج التالية:

نموذج I : $Y_i = \beta_1 + \beta_2 X_i + u_i$

نموذج II : $Y_i^* = \alpha_1 + \alpha_2 X_i^* + u_i$

حيث Y^* و X^* متغيرات قياسية، اثبت أن $\hat{\alpha}_2 = \hat{\beta}_2(S_x / S_y)$ وبالتالي وضح أنه على الرغم من أن معامل ميل الانحدار مستقل عن تغير نقطة الأصل، فإنه غير مستقل عن تغير وحدات القياس والأوزان.

6.6 اعتبر النماذج التالية:

$$\ln Y_i^* = \alpha_1 + \alpha_2 \ln X_i^* + u_i^*$$
$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

حيث $Y_i^* = w_1 Y_i$ و $X_i^* = w_2 X_i$، والـ w's ثابتة.

(a) وضح العلاقة بين المجموعتين السابقتين من معاملات الانحدار وأخطائهما القياسية.

(b) هل قيمة r^2 مختلفة في النموذجين السابقين؟

7.6 من بين الانحدارين (8.6.6) و (10.6.6)، أي منهما تفضل؟ ولماذا؟

8.6 بالنسبة للانحدار (8.6.6)، اختبر الفرض القائل بأن معامل الميل لا يختلف معنويًا عن 0.005.

9.6 من منحنى Phillips المعطى في (3.7.6)، هل من الممكن تقدير معدل البطالة الطبيعي؟ كيف؟

10.6 منحنى Engel للنفقات يربط بين إنفاق المستهلك على سلعة ما ودخله الكلي. افترض أن Y = الإنفاق الاستهلاكي على سلعة ما و X = دخل المستهلك، اعتبر النماذج التالية:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$
$$Y_i = \beta_1 + \beta_2 (1/X_i) + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 \ln X_i + u_i$$
$$\ln Y_i = \ln \beta_1 + \beta_2 (1/X_i) + u_i$$
$$Y_i = \beta_1 + \beta_2 \ln X_i + u_i$$

أي من هذه النماذج ستختار لمنحنى الإنفاق لـ Engel ولماذا؟ (**ملاحظة**: فسر معاملات الميل المختلفة، حدد أي منها مناسب للتعبير عن مرونة الإنفاق بالنسبة للدخل وهكذا).

11.6 اعتبر النموذج التالي:

$$Y_i = \frac{e^{\beta_1+\beta_2 X_i}}{1+e^{\beta_1+\beta_2 X_i}}$$

وفقًا للشكل السابق، هل هذا النموذج يعتبر نموذج انحدار خطي؟ واذا كانت إجابتك بلا، ما «الحيلة» الممكن اتباعها، إذا ودت هذه الحيل، لتحويل النموذج إلى نموذج انحدار خطي؟ وكيف يمكنك تفسير نتائج النموذج؟ ما هي الظروف التي يمكن أن تطبق قيمها مثل هذا النموذج؟

12.6 ارسم النماذج التالية (للتسهيل حذفنا الترميز i):

(a) $Y = \beta_1 X^{\beta_2}$, for $\beta_2 > 1, \beta_2 = 1, 0 < \beta_2 < 1, \ldots$

(b) $Y = \beta_1 e^{\beta_2 x}$, for $\beta_2 > 0$ and $\beta_2 < 0$

ناقش متى تكون هذه النماذج مناسبة.

مسائل: Problems

13.6 بافتراض ان لديك البيانات المعطاة في جدول (7.6) (*). وفق النموذج التالي بناء على هذه البيانات، واحصل على إحصاءات الانحدار التقليدية وفسر هذه النتائج:

$$\frac{100}{100-Y_i} = \beta_1 + \beta_2\left(\frac{1}{X_i}\right)$$

جدول (7.6)

Y_i	86	79	76	69	65	62	52	51	51	48
X_i	3	7	12	17	25	35	45	55	70	120

14.6 لقياس مرونة التبديل بين مدخل رأس المال ومدخل العمالة استخدم كلاً من Arrow، Chenery، Minhas و Solow (مؤلفون دالة الإنتاج المعروفة حاليًا باسم مرونة التبديل الثابتة) النموذج التالي (†):

$$\log\left(\frac{V}{L}\right) = \log\beta_1 + \beta_2 \log W + u$$

(*) المصدر Adapted from J. Johnston, Econometric Methods, 3d ed., McGraw-Hill, New York, 1984, p. 87. Actually this is taken from an econometric examination of Oxford University in 1975.

(†) Capital-Labor Substitution and Economic Efficiency," Review of Econmics and Statistics, August 1961, vol. 43, no. 5, pp. 225–254.

حيث (V/L) = القيمة المضافة لكل وحدة من العمالة .
L = مدخل العمالة .
W = معدل الأجر الحقيقي .

معامل β_2 يقيس مرونة التبديل بين العمالة ورأس المال (أي التغير النسبي في إنتاج المصنع / التغير النسبي في الأسعار النسبية) من البيانات المعطاة في جدول (8.6)، اثبت أن المرونة المقدرة تساوي 1.3338 وهذه القيمة لا تختلف معنوياً عن 1 .

15.6 جدول (9.6) يعطي بيانات عن GDP (إجمالي الناتج المحلي) بالنسبة للسلع المحلية و GDP للمستورد من سنغافورة خلال الفترة 1968 إلى 1982 . مؤشر GDP يُستخدم كمؤشر للتضخم بدلاً من CPI . سنغافورة تعتبر بلداً صغيراً ذا اقتصاد مفتوح، وتعتمد بشكل كبير على التجارة الخارجية .

جدول (8.6)

Industry	log(*V/L*)	log *W*
Wheat flour	3.6973	2.9617
Sugar	3.4795	2.8532
Paints and varnishes	4.0004	3.1158
Cement	3.6609	3.0371
Glass and glassware	3.2321	2.8727
Ceramics	3.3418	2.9745
Plywood	3.4308	2.8287
Cotton textiles	3.3158	3.0888
Woolen textiles	3.5062	3.0086
Jute textiles	3.2352	2.9680
Chemicals	3.8823	3.0909
Aluminum	3.7309	3.0881
Iron and steel	3.7716	3.2256
Bicycles	3.6601	3.1025
Sewing machines	3.7554	3.1354

المصدر: Damodar Gujarati, "A Test of AcmS Production Function: Indian Industries, 1958," Indian Journal of Industrial Relations, vol. 2, no. 1, July 1966. pp. 95–97.

جدول (9.6)

Year	GDP deflator for domestic goods, *Y*	GDP deflator for imports, *X*
1968	1000	1000
1969	1023	1042
1970	1040	1092
1971	1087	1105
1972	1146	1110
1973	1285	1257
1974	1485	1749
1975	1521	1770
1976	1543	1889
1977	1567	1974
1978	1592	2015
1979	1714	2260
1980	1841	2621
1981	1959	2777
1982	2033	2735

المصدر: Colin Simkin, "Does Money Matter in Singapore?" The Singapore Economic Review, vol. XXIX, no. 1, April 184 1984, Table 6, p.8.

لدراسة العلاقة بين الأسعار العالمية والمحلية، افترض أن لديك النماذج التالية:

1. $Y_t = \alpha_1 + \alpha_2 X_t + u_t$

2. $Y_t = \beta_2 X_t + u_t$

حيث Y = مؤشر GDP للسلع المحلية و X = GDP للواردات.

(a) كيف يمكنك الاختيار بين النموذجين السابقين؟

(b) قم بتطبيق كل من النموذجين السابقين على البيانات، وحدد أي منهما يعطي توفيقًا أدق للبيانات.

(c) ما هي النماذج الأخرى الممكن استخدامها للبيانات السابقة؟

16.6 بالعودة إلى البيانات المعطاة في تمرين (15.6) متوسطات Y و X هي كالتالي 1456 و 1760 على الترتيب، والأخطاء القياسية المرتبطة بهما هي 346 و 641.

قدر النموذج التالي:

$$Y_t^* = \alpha_1 + \alpha_2 X_t^* + u_t$$

حيث المتغيرات المنجمة هي متغيرات قياسية. فسر النتائج.

17.6 بالرجوع إلى جدول (3.6). أوجد معدل نمو الإنفاق على السلع المعمرة. ما هي شبه المرونة المقدرة؟ فسر نتائجك. هل من الممكن استخدام نموذج انحدار ثنائي اللوغاريتم لنموذج المتغير المنحدر عليه قيمة هو الإنفاق على السلع المعمرة والمتغير المنحدر هو الزمن؟ كيف يمكنك تفسير معامل الميل في هذه الحالة.

18.6 من البيانات المعطاة في جدول (3.6)، أوجد معدل نمو الإنفاق على السلع غير المعمرة، وقارن نتائجك مع تلك التي حصلت عليها في مسألة رقم (17.6).

19.6 ارجع إلى تمرين (7.1). والآن بعد أن تعرفت على العديد من أشكال الدوال، أي منها قد يكون مناسبًا لدراسة العلاقة بين الأثر العائد من الإعلان والمنفق عليه؟ وضح الخطوات الحسابية اللازمة لذلك.

APPENDIX ملحق A6

1.A6 اشتقاق مقدرات المربعات الصغرى في حالة الانحدار المار بنقطة الأصل
DERIVATION OF LEAST-SQUARES ESTIMATORS FOR REGRESSION THROUGH THE ORIGIN

نريد تصغير التالي:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_2 X_i)^2 \qquad (1)$$

بالنسبة لـ $\hat{\beta}_2$.

بتفاضل (1) بالنسبة لـ $\hat{\beta}_2$ **نحصل على**:

$$\frac{d \sum \hat{u}_i^2}{d\hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_2 X_i)(-X_i) \qquad (2)$$

وبمساواة (2) بالصفر والتبسيط **نحصل على**:

$$\hat{\beta}_2 = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (6.1.6) = 3$$

بالتعويض عن PRF: $Y_i = \beta_2 X_i + u_i$ في هذه المعادلة، نحصل على:

$$\hat{\beta}_2 = \frac{\sum X_i (\beta_2 X_i + u_i)}{\sum X_i^2}$$

$$= \beta_2 + \frac{\sum X_i u_i}{\sum X_i^2} \qquad (4)$$

[**لاحظ أن**: $E(\hat{\beta}_2) = \beta_2$]. وبالتالي:

$$E(\hat{\beta}_2 - \beta_2)^2 = E\left[\frac{\sum X_i u_i}{\sum X_i^2}\right]^2 \qquad (5)$$

إذا فصلنا الطرف الأيمن من المعادلة (5) مع ملاحظة أن X_i غير عشوائية و u_i ثابتة التباين، وغير مرتبطة نحصل على:

$$\text{var}(\hat{\beta}_2) = E(\hat{\beta}_2 - \beta_2)^2 = \frac{\sigma^2}{\sum X_i^2} \qquad (7.1.6) = (6)$$

لاحظ أنه من (2) وبعد المساواة بالصفر نحصل على:

$$\sum \hat{u}_i X_i = 0 \qquad (7)$$

من ملحق A3، الفقرة 1.A3، نرى أنه عندما يكون هناك جزء ثابت في النموذج، سيكون لدينا شرط إضافي لـ (7) وهو $\sum \hat{u}_i = 0$. من الخطوات الرياضية السابقة نرى

أن نموذج الانحدار المار بنقطة الأصل قد لا يكون مجموع الخطأ $\sum \hat{u}_i$، لا يساوي الصفر.

افترض أنك تريد فرض الشرط $\sum \hat{u}_i = 0$، في هذه الحالة لدينا التالي:

$$\begin{aligned} \sum Y_i &= \hat{\beta}_2 \sum X_i + \sum \hat{u}_i \\ &= \hat{\beta}_2 \sum X_i, \qquad \text{حيث } \sum \hat{u}_i = 0 \end{aligned} \tag{8}$$

وبذلك يكون لدينا:

$$\begin{aligned} \hat{\beta}_2 &= \frac{\sum Y_i}{\sum X_i} \\ &= \frac{\bar{Y}}{\bar{X}} = \frac{\text{متوسط } Y}{\text{متوسط } X} \end{aligned} \tag{9}$$

ولكن هذا المقدار ليس نفس المعطى في (3) أو (6.1.6) وبما أن $\hat{\beta}_2$ المعطاة في (3) غير متحيز (لماذا)، فإن $\hat{\beta}_2$ المعطى في (9) لن يكون غير متحيز.

وبالتالي، فالفكرة كالتالي في الانحدار المار بنقطة الأصل، لا يمكن أن يكون لدينا كل من $\sum \hat{u}_i X_i$ و $\sum \hat{u}_i$ يساويان الصفر كما في النموذج التقليدي. الشرط الوحيد المتحقق هو أن $\sum \hat{u}_i X_i$ يساوي الصفر.

تذكر أن:

$$Y_i = \hat{Y}_i + \hat{u} \tag{3.6.2}$$

وبإدخال الجمع على طرفي المعادلة السابقة والقسمة على N (حجم العينة) نحصل على:

$$\bar{Y} = \bar{\hat{Y}} + \bar{\hat{u}} \tag{10}$$

بما أنه في النموذج المار بنقطة الأصل $\sum \hat{u}_i$، وبالتالي $\bar{\hat{u}}$ قد لا تساوي الصفر، فإن لدينا التالي:

$$\bar{Y} \neq \bar{\hat{Y}} \tag{11}$$

أي أن متوسط قيم Y الفعلية قد لا يساوي متوسط قيم Y المقدرة. القيمتان متساويتان في حالة النموذج الذي يشمل الجزء الثابت، ويمكن التحقق من ذلك من (10.1.3).

سبق وذكرنا أنه في النموذج غير المشتمل على جزء ثابت، فإن r^2 قد تكون سالبة، أما في النموذج التقليدي لا يمكن أن تكون سالبة، هذا الشرط يمكن تفصيله كالتالي:

باستخدام (3.5.5a) يمكننا كتابة التالي:

$$r^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \qquad (12)$$

والآن في النموذج التقليدي أو النموذج الذي يشتمل على جزء ثابت. المعادلة (3.3.6) توضح التالي:

$$\text{RSS} = \sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2^2 \sum x_i^2 \leq \sum y_i^2 \qquad (13)$$

إلا إذا كانت $\hat{\beta}_2$ تساوي الصفر (أي أن X ليس لها تأثير عن Y). أي أنه في النموذج التقليدي يكون RSS≤TSS ولا يمكن أن تكون r^2 سالبة.

أما في النموذج الذي لا يشتمل على جزء ثابت، يمكن إثبات نفس الطريقة كالتالي:

$$\text{RSS} = \sum \hat{u}_i^2 = \sum Y_i^2 - \hat{\beta}_2^2 \sum X_i^2 \qquad (14)$$

(لاحظ أن: مجموع مربعات Y و X غير مصحح بالوسط الحسابي). الآن لا يوجد ما يضمن أن RSS سيكون دائمًا أقل من $\sum y_i^2 = \sum Y_i^2 - N\bar{Y}^2$ (الـ TSS) مما يعني أن RSS قد تكون أكبر من TSS مما يتطلب أن تكون r^2، بتعريفها التقليدي، سالبة. لاحظ أنه في مثل هذه الحالة تكون RSS أكبر من TSS إذا كان

$$\hat{\beta}_2^2 \sum X_i^2 < N\bar{Y}^2$$

2.A6 اثبات أن المتغيرات القياسية لها توقع يساوي الصفر، وتباين يساوي الوحدة
PROOF THAT A STANDARDIZED VARIABLE HAS ZERO MEAN AND UNIT VARIANCE

اعتبر المتغير العشوائي Y (r.v.) والذي له متوسط (عينة) $\bar{Y}$ وانحراف معياري (العينة) S_y، عرِّف التالي:

$$Y_i^* = \frac{Y_i - \bar{Y}}{S_y} \tag{15}$$

وبالتالي Y_i^* متغير قياسي. لاحظ أن القياسية تعني عملية مزدوجة: (1) تغير نقطة الأصل، وذلك موجود في بسط المعادلة (15) و(2) تغير المقياس، وذلك موجود في المقام. وبالتالي القياسية تعنى تغير في نقطة الأصل ووحدة القياس.

الآن

$$\bar{Y}_i^* = \frac{1}{S_y}\frac{\sum(Y_i - \bar{Y})}{n} = 0 \tag{16}$$

بما أن مجموع انحرافات المتغير عن وسطه الحسابي دائمًا تساوي الصفر. فإن متوسط المتغير القياسي هو الصفر (لاحظ أن: يمكننا إخراج S_y من التجميع، حيث إن قيمته معلومة).

الآن

$$\begin{aligned} S_{y^*}^2 &= \sum \frac{(Y_i - \bar{Y})^2/(n-1)}{S_y^2} \\ &= \frac{1}{(n-1)S_y^2}\sum(Y_i - \bar{Y})^2 \\ &= \frac{(n-1)S_y^2}{(n-1)S_y^2} = 1 \end{aligned}$$

لاحظ أن:

$$S_y^2 = \frac{\sum(Y_i - \bar{Y})^2}{n-1}$$

والذي يمثل تباين العينة لـ Y.

الفصل السابع

تحليل الانحدار المتعدد.. مشكلة التقدير

MULTIPLE REGRESSION ANALYSIS: THE PROBLEM OF ESTINATION

تمت دراسة النموذج الذي يشتمل على متغيرين اثنين فقط باستفاضة في الفصول السابقة، وإن كان هذا النموذج غير واقعي عمليًا. فمثلاً في مثالنا الخاص بالدخل والاستهلاك، تم الافتراض صراحةً بأن الدخل X يؤثر فقط على الاستهلاك Y. ولكن في إطار النظرية الاقتصادية الأمر ليس بهذه البساطة، فبالإضافة إلى الدخل، هناك العديد من المتغيرات المحتمل أن يكون لها تأثير على نفقات الاستهلاك.

وتعتبر ثروة المستهلك مثالاً واضح على هذه المتغيرات، وكمثال آخر الطلب على سلعة ما، حيث إنه من المحتمل ألا يعتمد فقط على سعر هذه السلعة الاجتماعية وهكذا. وبالتالي نحن نحتاج إلى امتداد لنموذج الانحدار البسيط ذي المتغيرين ليشمل نماذج تحتوي على أكثر من متغيرين اثنين فقط. إضافة متغيرات أكثر يجعلنا نتطرق إلى مناقشة نماذج الانحدار متعددة المتغيرات، أي النماذج التي يعتمد فيها المتغير التابع y على متغيرين مفسرين أو أكثر.

أبسط نموذج انحدار متعدد ممكن هو الانحدار ثلاثي المتغيرات ، والذي يحتوى على متغير تابع واحد ومتغيرين مفسرين. في هذا الفصل السابع والفصل الثامن سنقوم بدراسة تفصيلية لمثل هذا النموذج.

وعلى الرغم من أننا سنعتبر نماذج الانحدار الخطية متعددة المتغيرات، أي النماذج الخطية في المعلمات، إلا أنها قد تكون خطية أو غير خطية في المتغيرات.

1.7 النموذج ثلاثي المتغيرات.. الرموز والفروض: THE THREE - VARIABLE MODEL: NOTATION AND ASSUMPTIONS

لتقييم دالة انحدار المجتمع ثنائية المتغيرات (PRF) (2.4.2) يمكن أن نكتب PRF ثلاثية المتغيرات كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (1.1.7)$$

حيث Y هو المتغير التابع، X_2 و X_3 هما المتغيران المفسران (المنحدران)، u هو مقدار الخطأ العشوائي و i تعني المشاهدة رقم 1، في حالة وجود بيانات سلاسل زمنية، ترمز t إلى المشاهدة في الفترة الزمنية t.[1]

في المعادلة (1.1.7) β_1 يعبر عن الجزء المقطوع من المحور الصادي (الثابت). وهو يمثل متوسط التأثير على Y عندما تستبعد كل المتغيرات من النموذج، وبالتالي المعنى الفني له هو متوسط قيمة Y عندما تساوى X_2 وX_3 الصفر.

المعاملان β_2 و β_3 يسميان معاملات الانحدار الجزيئية، ومعناهما سيتم استعراضه لاحقًا.

وإذا استكملنا الشرح في إطار نموذج الانحدار الخطي التقليدي (CLRM) الذي قدمناه في الفصل 3، سنفترض التالي:

قيمة متوقعة تساوي الصفر لـ u_i أو

$$E(u_i \mid X_{2i}, X_{3i}) = 0 \qquad \text{for ench } i \qquad (2.1.7)$$

عدم وجود ارتباط تسلسلي أو

$$\text{cov}(u_i, u_j) = 0 \qquad i \neq j \qquad (3.1.7)$$

ثبات التباين

$$\text{var}(u_i) = \sigma^2 \qquad (4.1.7)$$

(1) لتماثل الترميز، المعادلة (1.1.7) يمكن كتابتها كالتالي:

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

مع اعتبار أن $X_{1i} = 1$ لكل i.

التغاير u_i وكل متغير X يساوي الصفر، أو

$$\text{cov}(u_i, X_{2i}) = \text{cov}(u_i, X_{3i}) = 0 \qquad (5.1.7)^{(2)}$$

لا يوجد تحيز في التوصيف، أو

(6.1.7) النموذج تم توصيفه بشكل صحيح

لا يوجد ارتباط خطي بين المتغيرات المفسرة X أو

(7.1.7) لا توجد علاقة خطية تامة بين X_2 وX_3.

وكما رأينا في الفصل (3)، فبالإضافة لذلك، نفترض أن نموذج الانحدار المتعدد هو نموذج خطي في المعلمات، أي أن قيم المتغيرات المفسرة (المنحدرة) ثابتة في العينات المتكررة، وأن هناك تباينًا كما في قيم المتغيرات المنحدرة.

المنطق وراء الفروض (2.1.7) إلى (6.1.7) هو نفس المنطق الذي ناقشناه في الفقرة 2.3. الفرض (7.1.7)، القائل بأنه لا توجد علاقة خطية تامة بين X_2 و X_3، يتم التعبير عنه فنيًا بأنه فرض عدم وجود ارتباط خطي أو ارتباط خطي متعدد إذا كان هناك أكثر من علاقة خطية تامة موجودة حديثًا وتحتاج لتفسير.

وبشكل عام، عدم وجود ارتباط خطي يعني أنه لا يمكن كتابة أحد المتغيرات المنحدرة في صورة توليفة خطية من المتغيرات المنحدرة الأخرى الموجودة في النموذج.

وبشكل أكثر دقة، عدم وجود ارتباط خطي يعني أنه لا توجد فئة من الفئات X_2 و X_3 بحيث إنهما معًا يساويان الصفر ويحققان التالي:

$$\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0 \qquad (7.1.8)$$

إذا وجدت مثل هذه العلاقة الخطية التامة، فإن X_2 و X_3 يقال إنهما مرتبطان خطيًا أو تابعان خطيًا، على الجانب الآخر، إذا تحقق (8.1.7) فقط عندما يكون $\lambda_2 = \lambda_3 = 0$ فإن X_2 و X_3 يقال إنهما مستقلان خطيًا، وبالتالي إذا كان

$$X_{2i} = -4\,X_{3i} \quad \text{أو} \quad X_{2i} + 4\,X_{3i} = 0 \qquad (9.1.7)$$

(2) هذا الغرض يتحقق تلقائيًا إذا كان X_2 و X_3 متغيرين غير عشوائيين مع تحقق (2.1.7).

فإن المتغيرين تابعان خطيًا، وإذا وجد الاثنان معًا في نموذج الانحدار سيكون لدينا ارتباط خطي تام، أو علاقة خطية تامة بين متغيرين اثنين منحدرين.

وعلى الرغم من أننا سنتناول بالتفصيل مشكلة الارتباط الخطي المتعدد في الفصل 10، فإنه ليس من الصعب استعراض الآن المنطق وراء فرض عدم وجود ارتباط خطي بين المتغيرات المفسرة. فافترض أنه في (1.1.7) Y، X_2 و X_3 تمثل نفقات الاستهلاك، الدخل وثروة المستهلك بالترتيب.

بافتراض أن نفقات الاستهلاك مرتبطة خطيًا مع الدخل والثروة، فإن النظرية الاقتصادية تفترض أن الثروة والدخل لهما بعض التأثير المستقل على الاستهلاك، وإذا لم يوجد ذلك فإنه من غير المنطقي اعتبار كل من الدخل والثروة كمتغيرات مفسرة في النموذج. وعلى الجانب الآخر، إذا وجدت علاقة خطية تامة بين الدخل والثروة، فإن لدينا فعليًا متغير مستقل واحد وليس اثنين وليس من الممكن افتراض التأثير المنفصل للدخل وللثروة على الاستهلاك. لرؤية ذلك بوضوح، دع $X_{3i} = 2X_{2i}$ في انحدار الاستهلاك - الدخل - الثروة. بالتالى فالانحدار (1.1.7) يصبح كالتالي:

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + \beta_3 (2X_{2i}) + u_i \\ &= \beta_1 + (\beta_2 + 2\beta_3) X_{2i} + u_i \qquad (10.1.7) \\ &= \beta_1 + \alpha X_{2i} + u_i \end{aligned}$$

حيث $\alpha = (\beta_2 + 2\beta_3)$. أي أننا في الحقيقة لدينا متغيران اثنان فقط وليس ثلاثة متغيرات في هذا الانحدار السابق. والأكثر من ذلك إذا طبقنا انحدار (10.1.7) وحصلنا على α لن يكون من الممكن تقديرا لتأثير المنفصل لـ $X_2 (=\beta_2)$ ولـ $X_3 (=\beta_3)$ على Y حيث إن α تعطي التأثير المشترك لـ X_2 و X_3 على Y.[3]

باختصار، فإن فرض عدم وجود ارتباط خطي يتطلب أن PRF يحتوي فقط على المتغيرات التي لا يمكن التعبير عنها في صورة دوال خطية تامة من متغير أو أكثر من المتغيرات الأخرى الموجودة في النموذج. وعلى الرغم من أننا سنتناول هذه النقطة بمزيد من التفاصيل في الفصل 10، إلا أن بعض النقاط المهمة يجب ملاحظتها هنا:

(3) بشكل رياضي فإن $\alpha = (\beta_2 + \beta_3)$ هي معادلة واحدة في مجهولين اثنين ولا يوجد حل وحيد لتقدير β_2 و β_3 من قيمة α المقدرة.

أولاً، فرض عدم وجود ارتباط خطي مهم بالنسبة للجانب النظري للنموذج (PRF). في الواقع العملي، عندما نقوم بتجميع بيانات لتحليل تجريبي لا يوجد ما يضمن أنه لن يكون هناك ارتباط بين المتغيرات المنحدرة، بل في واقع الأمر، فإنه في غالبية العمل التطبيقي يكون من المستحيل الحصول على متغيرين أو أكثر (اقتصاديين) غير مرتبطين نوعًا ما.

كما سنرى في مثالنا التوضيحي الذي سيأتي لاحقًا في الفصلين السابع والثامن، ما نرجوه هو عدم وجود علاقة تامة فقط بين المتغيرات المنحدرة كما في المعادلة (9.1.7).

ثانيًا، ضع في الاعتبار أننا نتحدث فقط على علاقة خطية تامة بين متغيرين أو أكثر. الارتباط الخطي المتعدد لا يفسد العلاقة غير الخطية بين المتغيرات. افترض أن $X_{3i} = X_{2i}^2$ فهذا لا يخالف فرض عدم وجود ارتباط خطي تام، حيث إن العلاقة بين المتغيرين علاقة غير خطية.

2.7 تفسير معادلة الانحدار المتعدد :
INTERPRETATION OF MULTIPLE REGRESSION EQUATION

وفقًا لفروض نموذج الانحدار التقليدي، فإن التوقع الشرطي لـ y لكل من طرفي المعادلة (1.1.7) تحصل على:

$$E(Y_i \mid X_{2i}, X_{3i}) = \beta_1 + \beta_2 X_{2i} + \beta_{3i} X_{3i} \qquad (1.2.7)$$

بمعنى آخر، (1.2.7) تعطي التوقع الشرطي أو القيمة المتوقعة لـ y وفقًا لقيم محددة معطاة لكل من X_2، X_3. وبالتالي، في حالة وجود متغيرين اثنين، تحليل الانحدار المتعدد هو تحليل انحدار مشروط بقيم محددة للمتغيرات المنحدرة ونحصل على القيمة المتوسطة لـ Y أو متوسط الاستجابة لـ Y وفقًا للقيم المعطاة للمتغيرات المنحدرة.

3.7 مغزى معاملات الانحدار الجزيئية :
THE MEANING OF PARTIAL REGRESSION COEFFICIENTS

كما سبق وذكرنا، معاملات الانحدار β_2 و β_3 معروفة باسم معاملات الانحدار الجزيئية أو معاملات الميل الجزيئية. المعنى المقصود من معامل الانحدار الجزيئي هو كالتالي: β_2 تعني التغير في متوسط Y، $E(Y)$، بالنسبة للتغير بمقدار الوحدة في X_2، مع افتراض ثبات قيمة X_3. ولكن كتفاضل جزيئي، فإنها تعطي التأثير «المباشر» أو «الصافي» للتغير بمقدار الوحدة في X_2 على متوسط قيمة Y، وهذا التأثير صافي أي

بعيد عن تأثير X_3 على متوسط Y. وبالمثل فإن β_3 تعيش التغير في القيمة المتوسطة لـ Y لكل وحدة تغير في X_3، بافتراض ثبات X_2.[4] أي أنه يعطي التأثير «المباشر» أو «الصافي» للتغير في X_3 على القيمة المتوسطة لـ Y، بعيداً عن أي تأثير لـ X_2 على متوسط Y.[5]

كيف يمكن فعليًا تثبيت تأثير أحد المتغيرات المنحدرة؟ لشرح ذلك، دعنا نستخدم مرة أخرى مثال وفيات الأطفال. نذكر في هذا المثال كان: Y = وفيات الأطفال (CH)، X_2 = PGNP، X_3 = معدل تعليم المرأة (FLR). دعنا نفترض أننا نريد تثبيت أثر FLR. بما أن FLR ممكن يكون له تأثير على CM مثله مثل الـ PGNP ما يمكن عمله هو حذف التأثير الخطي لـ FLR من كل من CM و PGNP عن طريق عمل انحدار لـ CM على FLR مرة وانحدار لـ PGNP على FLR مرة أخرى، ثم نرى البواقي التي نحصل عليها من هذه الانحدارات. باستخدام البيانات المعطاة في جدول (4.6)، نحصل على الانحدارات التالية:

$$\begin{aligned} \text{CM}_i &= 263.8635 \;-\; 2.3905\ \text{FLR}_i + \hat{u}_{1i} \\ \text{se} &= (12.2249) \quad (0.2133) \qquad\qquad r^2 = 0.6695 \end{aligned} \tag{1.3.7}$$

حيث $\hat{u}_{1i}$ تمثل البواقي لهذا الانحدار.

$$\begin{aligned} \text{PGNP}_i &= -39.3033 \;+\; 28.1427\ \text{FLR}_i + \hat{u}_{2i} \\ \text{se} &= (734.9526) \quad (12.8211) \qquad\qquad r^2 = 0.0721 \end{aligned} \tag{2.3.7}$$

حيث $\hat{u}_{2i}$ تمثل بواقي هذا الانحدار والآن:

$$\hat{u}_{1i} = (\text{CM}_i - 263.8635 + 2.3905\ \text{FLR}_i) \tag{3.3.7}$$

يمثل الجزء من CM الباقي بعد حذف التأثير (الخطي) لـ FLR. بالمثل،

$$\hat{u}_{2i} = (\text{PGNP}_i + 39.3033 - 28.1427\ \text{FLR}_i) \tag{4.3.7}$$

يمثل الجزء من PGNP الباقي بعد حذف التأثير الخطي لـ RLR.

وبالتالي، إذا قمنا الآن بعمل انحدار لـ $\hat{u}_{1i}$ على $\hat{u}_{2i}$، والذي يعتبر «منقى» من التأثير الخطي لـ FLR، هل سنحصل على التأثير الصافي لـ PGNP على CM؟ هذا بالضبط ما سنحصلْ عليه (انظر ملحق A7، الفقرة 2.A7). نتائج الانحدار كالتالي:

---

(4) للقارئ ذي العقلية الحسابية فإن β_2 و β_3 هما تفاضلان جزيئيان لـ $E(Y \mid X_2, X_3)$ بالنسبة لـ β_2 و β_3.

(5) أحيانًا سيتم استخدام المصطلحات التالية بشكل تبادلي وهي: افتراض الثبات، التحكم في، أو السماح بتأثير، تصحيح التأثير وتوضيح التأثير.

$$\hat{u}_{1i} = -0.0056\hat{u}_{2i} \tag{5.3.7}$$
$$\text{se} = \quad (0.0019) \qquad r^2 = 0.1152$$

لاحظ أن: هذا الانحــدار لا يوجـد فـيـه جـزء ثابـت (الجـزء المقطوع من المحـور الصادي)، حيث إن متوسط بواقي الـ OLS $\hat{u}_{1i}$ و $\hat{u}_{2i}$ يساوي الصفر (لماذا؟).

الآن معامل الميل المساوي 0.0056 − يعني التأثير «الحقيقي» أو الصافي لتغير وحدة في PGNP على CM أو الميل الحـقيـقي لـ CM بالنسـبة لـ PGNP أي يسـاوي مـعـامل الانحدار الجزئي لـ CM بالنسبة لـ PGNP هو β_2.

القراء الذين يرغبون في الحصول على معامل الانحدار الجزئي لـ CM بالنسبة لـ FLR يمكن أن يتـبـعـوا نفس الخطوات السـابقـة أولاً بعـمـل انحـدار لـ CM على PGNP ويحـصلون على بواقي هذا الانـحـدار ($\hat{u}_{1i}$) ثم انحـدار FLR على PGNP ويحـصلون على بواقي هذا الانحدار أيضًا ($\hat{u}_{2i}$) ثم يقومون بعمل انحدار لـ $\hat{u}_{1i}$ على $\hat{u}_{2i}$.

وأعتقد أن الفكرة أصبحت واضحة الآن.

هل مطلوب القيام بكل هذه الخطوات العديدة في كل مرة نريد فيها الحصول على معامل الانحدار الجزئي الحقيقي؟ لحسن الحظ ليس بالضرورة القيام بكل ذلك، حيث إن طريقة OLS تقوم بعمل ذلك بشكل سريع وواضح نسبيًا وسيتم شرحها في الفقرة التالية. فالخطوات المتعددة السابقة تم استعراضها لتوضيح الفكرة العامة، وتوضيح المقصود بمعامل انحدار جزئي.

4.7 مقدرات OLS و ML لمعاملات الانحدار الجزيئية :
OLS AND ML ESTIMATION OF THE PARTIAL REGRESSION COEFFICIENTS

لتقدير معلمات نموذج انحدار ثلاثي المتغيرات (1.1.7)، سنستعرض أولاً طريقة المربعات الصغرى العادية (OLS) التي قدمناها في الفصل (3) وبعد ذلك سنستعرض سريعًا لطريقة الإمكان الأعظم (ML) التي قدمناها في الفصل (4).

مقدرات OLS : **OLS Estimators**

لإيجاد مقدرات OLS، دعنا أولاً نكتب دالة انحدار العينة (SRF) الخاصة بـ PRF في (1.1.7) كالتالي:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \quad (1.4.7)$$

حيث $\hat{u}_i$ هو مقدار البواقي، المناظر من العينة لمقدار الخطأ العشوائي u_i.

كما سبق وذكرنا في الفصل (3)، فإن التقدير بطريقة OLS يعتمد على اختيار قيم المعلمات المجهولة، بحيث إن مجموع مربعات البواقي $\sum \hat{u}_i^2$(RSS) يكون أصغر ما يمكن، وباستخدام الرموز، فإن ذلك يعني التالي:

$$\min \sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \quad (2.4.7)$$

حيث يتم الحصول على صيانة الـ RSS بعمل بعض الخطوات الجبرية البسيطة على المعادلة (1.4.7).

أبسط الطرق المباشرة للحصول على المقدرات من خلال تصغير (2.4.7) تتم من خلال الحصول على تفاضل الدالة بالنسبة للمجاهيل الموجودة ومساواة ذلك بالصفر وحل هذه المعادلات آنيًا. وكما هو موضح في ملحق A7 الفقرة 1.A7 فإن هذه الطريقة تؤدي إلى المعادلات الطبيعية التالية [انظر معادلات (4.1.3) و (5.1.3)]:

$$\bar{Y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{X}_2 + \hat{\beta}_3 \bar{X}_3 \quad (3.4.7)$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} \quad (4.4.7)$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 \quad (5.4.7)$$

من المعادلة (3.4.7) نجد أن

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3 \quad (6.4.7)$$

وهذا هو مقدر OLS للجزء الثابت الخاص با لمجتمع β_1.

وباستخدام الحروف الصغيرة للتعبير عن الانحرافات عن قيم العينة المتوسطة، يمكننا اشتقاق الصيغ التالية من المعادلات الصلبة الموجودة في المعادلات (3.4.7) إلى (5.4.7).

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \quad (7.4.7)^{(6)}$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i} x_{3i}\right)^2} \quad (8.4.7)$$

---

(6) هذا المقدر مساو لـ (5.4.7) كما هو موضح في الملحق A7، فقرة 2.A7.

وذلك يمثل مقدرات OLS لمعاملات انحدار المجتمع الجزيئية لـ β_2 و β_3 بالترتيب.

وعمومًا لاحظ التالي: (1) المعادلان (7.4.7) و (7.4.8) متماثلان بطبيعتهما، حيث يمكن الحصول على أحدهما من الآخر بتغيير أماكن X_2 و X_3، (2) المقام في هاتين المعادلتين واحد، (3) حالة ثلاثة متغيرات هي امتداد طبيعي لحالة متغيرين اثنين.

تباين مقدرات OLS وأخطاؤها القياسية:

Variances and Standard errors of OLS estimations

بعد الحصول على مقدرات OLS لمعاملات الانحدار الجزيئية، يمكن الحصول على تباينات، والأخطاء القياسية لهذه المقدرات بالطريقة المذكورة في ملحق 3.A3. وكما في حالة وجود متغيرين اثنين فقط، فإننا نحتاج إلى الأخطاء القياسية لسببين رئيسيين: لتكوين فترات الثقة والقيام باختبارات الفروض. هذه الصيغ كالتالي: [7]

$$\operatorname{var}(\hat{\beta}_1) = \left[\frac{1}{n} + \frac{\bar{X}_2^2 \sum x_{3i}^2 + \bar{X}_3^2 \sum x_{2i}^2 - 2\bar{X}_2\bar{X}_3 \sum x_{2i}x_{3i}}{\sum x_{2i}^2 \sum x_{3i}^2 - \left(\sum x_{2i}x_{3i}\right)^2}\right] \cdot \sigma^2 \qquad (9.4.7)$$

$$\operatorname{se}(\hat{\beta}_1) = +\sqrt{\operatorname{var}(\hat{\beta}_1)} \qquad (10.4.7)$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sum x_{3i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \qquad (11.4.7)$$

وذلك مساو لـ:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2\left(1 - r_{23}^2\right)} \qquad (12.4.7)$$

حيث r_{23} هو معامل ارتباط العينة بين X_2 و X_3 كما سبق تعريفه في الفصل (3). [8]

$$\operatorname{se}(\hat{\beta}_2) = +\sqrt{\operatorname{var}(\hat{\beta}_2)} \qquad (13.4.7)$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sum x_{2i}^2}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2}\sigma^2 \qquad (14.4.7)$$

---

(7) هذه الاشتقاقات يتم الحصول عليها بشكل أسهل من خلال استخدام المصفوفات. القارئ الذي يريد التعمق في ذلك يرجع إلى ملحق C.

(8) باستخدام تعريف r المعطي في الفصل 3، لدينا

$$r_{23}^2 = \frac{\left(\sum x_{2i}x_{3i}\right)^2}{\sum x_{2t}^2 \sum x_{3t}^2}$$

أو بالمثل :

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)} \tag{15.4.7}$$

$$\operatorname{se}(\hat{\beta}_3) = +\sqrt{\operatorname{var}(\hat{\beta}_3)} \tag{16.4.7}$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2}\sqrt{\sum x_{3i}^2}} \tag{17.4.7}$$

في كل هذه الصيغ σ^2 هو التباين (الثابت) لمجتمع الأخطاء u_i.

وباتباع خطوات ملحق A3، الفقرة 5.A3، يمكن للقارئ أن يثبت أن المقدر غير المتحيز لـ σ^2 يأخذ الشكل التالي :

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3} \tag{18.4.7}$$

لاحظ التماثل بين مقدر σ^2 الحالي ومقدره في حالة وجود متغيرين اثنين فقط $\left[\hat{\sigma}^2 = \left(\sum \hat{u}_i^2\right)/(n-2)\right]$. درجات الحرية الآن هي $(n-3)$ حيث إننا عند تقدير $\sum \hat{u}_i^2$ نحتاج أولاً لتقدير β_1، β_2، β_3 مما يستهلك 3 درجات حرية (هذه نتيجة عامة، أي أنه في حالة وجود أربعة متغيرات، فإن درجات الحرية ستكون $n-4$).

مقدار $\hat{\sigma}^2$ يمكن حسابه من (18.4.7) بمجرد الحصول على البواقي، ولكن يمكن الحصول عليها ببساطة باستخدام العلاقة التالية (للإثبات انظر الملحق A7، الفقرة 3.A7) :

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i} \tag{19.4.7}$$

وهذه هي الصيغة الخاصة بثلاثة متغيرات والمناظرة للعلامة الموجودة في (6.3.3).

خصائص مقدرات OLS : Properties of OLS Estimators

خصائص مقدرات OLS في حالة نموذج الانحدار المتعدد مناظره لمثيلتها في حالة النموذج ثنائي المتغيرات. وذلك بالأخص في التالي :

1 - خط الانحدار ثلاثي المتغيرات (المساحة) يمرر من خلال المتوسطات $\bar{Y}$ ، $\bar{X}$ و $\bar{X}_1$ ، وذلك مثبت من خلال (3.4.7) [ارجع إلى المعادلة (17.3) في النموذج ثنائي المتغيرات]. هذه الخاصية عامة. وبالتالي في حالة نموذج انحدار فيه k متغير [متغير منحدر عليه و $(k-1)$ متغيرات منحدرة] فإن :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{20.4.7}$$

فإن لدينا :

$$\hat{\beta}_1 = \bar{Y} - \beta_2 \bar{X}_2 - \beta_3 \hat{X}_3 - \cdots - \beta_k \bar{X}_k \tag{21.4.7}$$

2 – القيمة المتوسطة للمقدار Y_i $(\hat{Y}_i =)$ تساوي القيمة المتوسطة للقيم الحقيقية Y_i . وهذا من السهل إثباته :

$$\begin{aligned} \hat{Y}_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \\ &= (\bar{Y} - \hat{\beta}_2 \bar{X}_2 - \hat{\beta}_3 \bar{X}_3) + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} \quad (\text{لماذا؟}) \\ &= \bar{Y} + \hat{\beta}_2 (X_{2i} - \bar{X}_2) + \hat{\beta}_3 (X_{3i} - \bar{X}_3) \\ &= \bar{Y} + \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \end{aligned} \tag{22.4.7}$$

وكالمعتاد، تستخدم الحروف الصغيرة للتعبير عن الانحرافات عن متوسطتها المناظرة.

بجمع طرفي المعادلة (22.4.7) على قيم العينة وقيمتها على حجم العينة n نحصل على $\bar{\hat{Y}} = \bar{Y}$ (لاحظ أن : $\sum x_{2i} = \sum x_{3i} = \sum x_{3i} = 0$. لماذا؟)

لاحظ أنه باستخدام ميزة المعادلة (22.4.7) يمكننا كتابة التالي :

$$\hat{y}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} \tag{23.4.7}$$

حيث $\hat{y}_i = (\hat{Y}_i - \bar{Y})$

وبالتالي الـ SRF للمعادلة (1.4.7) يمكن كتابته في صورة انحرافات كالتالي :

$$y_i = \hat{y}_i + \hat{u}_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \tag{24.4.7}$$

3 – $\sum \hat{u}_i = \bar{\hat{u}} = 0$ ، وذلك يمكن إثباته من (24.4.7) [استخدم جمع طرفي المعادلة (24.4.7) على قيم العينة].

4 – البواقي $\hat{u}_i$ غير مرتبطة مع X_{2i} و X_{3i} ، أي أن، $\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i}$

(الإثبات موجود في ملحق 1.A7).

5 – البواقي $\hat{u}_i$ غير مرتبطة مع $\hat{Y}_i$ ، أي أن $\sum \hat{u}_i \hat{Y}_i = 0$. لماذا؟

[اضرب طرفي المعادلة (23.4.7) في $\hat{u}_i$ وجمع على قيم العينة].

6 – من (12.4.7) و (15.4.7) مثبت أن r_{23}، معامل الارتباط بين X_2 و X_3، يزداد في اتجاه 1، تباين $\hat{\beta}_2$ و $\hat{\beta}_3$ يزداد للقيمة المعطاة لـ σ^2 و $\sum x_{2i}^2$ أو $\sum x_{3i}^2$. في النهايات، عندما $r_{23} = 1$ (أي ارتباط تام) فإن التباينات تكون ما لا نهاية. المعنى المرتبط بذلك سيتم

استعراضه بالتفصيل في الفصل (10)، ولكن مبدئيًا يمكن للقارئ أن يرى أنه مع زيادة r_{23} يكون من الصعب معرفة القيم الحقيقية لـ β_2 و β_3. [المزيد من التفاصيل عن هذا الموضوع موجودة في الفصل التالي، ولكن ارجع إلى المعادلة (10.1.7) لأخذ صورة عامة عن هذه النقطة].

7 - من الواضح أيضًا من (12.4.7) و (15.4.7) أنه بالنسبة لقيم محددة لـ r_{23} و $\sum x_{2i}^2$ أو $\sum x_{3i}^2$ ، فإن تباينات مقدرات OLS هي متناسبة مباشرة مع σ^2 أي أنها تتزايد مع زيادة σ^2. وبالمثل بالنسبة لقيم محددة لـ σ^2 و r_{23}، تباين $\hat{\beta}_2$ متناسب عكسيًا مع $\sum x_{2i}^2$ أي أنه كلما زاد التباين في قيم العينة الخاصة بـ X_2 كلما قل تباين $\hat{\beta}_2$ ، وبالتالي يتم تقدير β_2 بدقة أكبر. ويمكن الوصول لنتيجة مماثلة خاصة بتباين $\hat{\beta}_3$.

8 - وفقًا لفروض نموذج الانحدار الخطي التقليدية، والتي تم التعرض لها تفصيلاً في الفقرة 1.7 ، يمكن إثبات أن مقدرات OLS لمعاملات الانحدار الجزيئية ليست فقط خطية وغير متحيزة، ولكن أيضًا أقل تباينًا داخل فئة المقدرات الخطية غير المتحيزة، وباختصار، فإن ذلك يعني أن هذه المقدرات BLUE : أو بشكل مماثل، فإنها تحقق نظرية Gauss-Markov . (الإثبات موجود في ملحق A3، فقرة 6.A3 لحالة متغيرين اثنين فقط، وسيتم استعراض ذلك بشكل مختصر في صورة مصفوفات في ملحق 2).

مقدرات الإمكان الأعظم : Maximum Likelihood Estimators

لاحظنا في الفصل (4)، أنه بافتراض أن u_i، خطأ المجتمع، يتبع التوزيع الطبيعي بتوقع يساوي الصفر وتباين ثابت σ^2، فإن مقدرات الإمكان الأعظم (ML) ومقدرات OLS لمعاملات الانحدار في حالة وجود متغيرين اثنين متساويين. هذا التساوي يتحقق أيضًا في النماذج التي تحتوي على أي عدد من المتغيرات (للإثبات انظر الملحق A7، الفقرة 4.A7) وعمومًا هذا غير صحيح بالنسبة لمقدر σ^2. فمن الممكن إثبات أن مقدر ML لـ σ^2 هو $\sum \hat{u}_i^2/n$ بغض النظر عن عدد المتغيرات الموجودة في النموذج، في حين أن مقدر OLS لـ σ^2 هو $\sum \hat{u}_i^2/(n-2)$ في حالة وجود متغيرين اثنين ويساوي $\sum \hat{u}_i^2/(n-3)$ في حالة وجود ثلاثة متغيرات و $\sum \hat{u}_i^2/(n-k)$ في حالة وجود k متغير كما في المعادلة (20.4.7). باختصار فإن مقدار OLS لـ σ^2 تأخذ في

اعتبارها عدد درجات الحرية، في حين أن مقدر ML لـ σ^2 لا يهتم بذلك. بالطبع إذا كانت n كبيرة جداً، فإن مقدرات ML و OLS لـ σ^2 ستكون تقريباً متساوية. (لماذا؟)

5.7 معامل التحديد المتعدد R^2.. ومعامل الارتباط المتعدد R :
THE MULTIPLE COEFFICIENT OF DETERMINATION R^2 AND THE MULTIPLE COEFFICIENT OF CORRELATION R

في حالة وجود متغيرين اثنين فقط في نموذج الانحدار رأينا أن r^2 المعرفة في (5.5.3) تقيس جودة توفيق معادلة الانحدار، أي أنه يمثل نسبة التباين الكلي في المتغير التابع Y والتي يمكن تفسيرها من خلال المتغير المفسر (الوحيد) X. هذا المفهوم الخاص بـ r^2 يمكن بسهولة تعميمه على نماذج الانحدار التي تشتمل على أكثر من متغيرين. وبالتالي في حالة وجود ثلاثة متغيرات ستكون مهتمين بمعرفة نسبة تباين Y المفسرة من خلال X_2 و X_3 معاً. هذه المعلومة نحصل عليها من خلال ما يسمى معامل التحديد المتعدد، ويرمز له بالرمز R^2 وهو مناظر لمفهوم r^2. للحصول على R^2، سنتبع طريقة اشتقاق r^2 المعطاة في فقرة 5.3. كالتالي:

$$\begin{aligned} Y_i &= \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \\ &= \hat{Y}_i + \hat{u}_i \end{aligned} \tag{1.5.7}$$

حيث $\hat{Y}_i$ هي القيمة المقدرة لـ Y_i من معادلة الانحدار، وهي مقدر للقيمة الحقيقية $E(Y_i \mid X_{2i}X_{3i})$. وباستخدام الحروف الصغيرة للتغير عن الانحرافات عن القيم المتوسطة فإن المعادلة (1.5.7) يمكن كتابتها كالتالي:

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \\ &= \hat{y}_i + \hat{u}_i \end{aligned} \tag{2.5.7}$$

بتربيع طرفي المعادلة (2.5.7) والتجميع على رقم العينة، نحصل على:

$$\begin{aligned} \sum y_i^2 &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 + 2\sum \hat{y}_i \hat{u}_i \\ &= \sum \hat{y}_i^2 + \sum \hat{u}_i^2 \qquad \text{(Why?)} \quad \text{(لماذا؟)} \end{aligned} \tag{3.5.7}$$

وبالصيغ الكلامية، فإن المعادلة (3.5.7) تعبر عن مجموع المربعات الكلي (TSS) المساوي لمجموع المربعات المفسرة (ESS) + مجموع مربعات البواقي (RSS). وبالتعويض عن قيمة $\sum \hat{u}_i^2$ من (19.4.7) في المعادلة السابقة، نحصل على:

$$\sum y_i^2 = \sum \hat{y}_i^2 + \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

وبإعادة ترتيب هذه المعادلة، نحصل على:

$$\text{ESS} = \sum \hat{y}_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} \quad \textbf{(4.5.7)}$$

والآن، بالتعريف فإن:

$$R^2 = \frac{\text{ESS}}{\text{TSS}}$$

$$= \frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{\sum y_i^2} \quad \textbf{(5.5.7)}^{(9)}$$

[انظر لـ (5.5.7) و (6.5.3)].

حيث إن الكميات المكونة لـ (5.5.7) يمكن حسابها بشكل روتيني، فإن R^2 يمكن بسهولة الحصول عليه. لاحظ أن R^2، مثل r^2، يقع بين 0 و 1. إذا كان مساويًا لـ 1، فإن ذلك يعني أن معادلة الانحدار تفسر 100% من تباين Y. وعلى الجانب الآخر، إذا كان يساوي 0 فإن النموذج لا يفسر أي شيء من تباين Y. وعمومًا فإن R^2 تقع بين هاتين القيمتين. توفيق النموذج يقال إنه أفضل كلما اقتربت R^2 إلى 1.

تذكر أنه في حالة وجود متغيرين اثنين فقط، فإننا نعرف القيمة r على أنها معامل الارتباط، وهي مؤشر لقياس درجة الارتباط (الخطي) بين متغيرين.

أما إذا كان لدينا ثلاثة متغيرات أو أكثر، فإننا نستخدم معامل الارتباط المتعدد ويرمز له بالرمز R، وهو يقيس درجة الارتباط بين Y وكل المتغيرات المفسرة معًا. وعلى الرغم من أن r ممكن أن يكون موجبًا أو سالبًا فإن R دائمًا موجب. في الواقع العملي، عمومًا R لها أهمية أقل. القيمة ذات المعنى الأهم هي R^2.

قبل التعمق في التفاصيل، دعنا أولاً نلاحظ العلاقة التالية بين R^2 وتباين معاملات الانحدار الجزيئية في حالة نموذج انحدار له k متغير كما في (20.4.7):

$$\text{var}\,(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \quad \textbf{(6.5.7)}$$

---

(9) لاحظ أن R^2 يمكن حسابه أيضًا كالتالي:

$$R^2 = 1 - \frac{\text{RSS}}{\text{TSS}} = 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} = 1 - \frac{(n-3)\hat{\sigma}^2}{(n-1)S_y^2}$$

حيث $\hat{\beta}_j$ هي معامل الانحدار الجزئية للمتغير المنحدر X_j وR_j^2 هي R^2 في انحدار X_j على $(k-2)$ من المتغيرات المنحدرة الباقية. [لاحظ: هناك $(k-1)$ من المتغيرات المنحدرة في نموذج انحدار ذي k متغير]. وعلى الرغم من أن أهمية المعادلة (6.5.7) ستظهر بوضوح في الفصل (10) عند استعراض موضوع الارتباط المتعدد، إلا أنه يمكن الآن ملاحظة أن هذه الصيغة ما هي إلا تعميم للصيغة (12.4.7) أو (15.4.7) لنموذج الانحدار ثلاثي المتغيرات، متغير منحدر عليه واحد ومتغيران منحدران.

6.7 مثال 1.7: وفيات الأطفال وعلاقتها بـ PGNP ومعدل تعلم القراءة والكتابة للمرأة: EXAMPLE 7.1: CHILD MORTALITY IN RELATION TO PER CAPITA GNP AND FEMALE LITERACY RATE

في الفصل (6)، استعرضنا وفيات الأطفال (CM) وعلاقتها بالـ (PGNP). ووجدنا أن PGNP لها تأثير سلبي على CM. كما هو متوقع. والآن دعنا نستخدم تعلم القراءة والكتابة للمرأة المقاس بمعدل تعلم القراءة والكتابة للمرأة (FLR). مبدئيًا نتوقع أن FLR سيكون له أيضًا تأثير سلبي على CM. والآن دعنا نستخدم هذين المتغيرين معًا في النموذج، حيث نريد الوصول إلى التأثير الخاص بكل متغير من المتغيرين المفسرين السابقين. أي أننا نريد تقدير معامل الانحدار (الجزئي) لكل متغير منحدر. وبالتالي فإن نموذجنا كالتالي:

$$\text{CM}_i = \beta_1 + \beta_2 \text{PGNP}_i + \beta_3 \text{FLR}_i + u_i \qquad (1.6.7)$$

البيانات الرئيسية موجودة في جدول (4.6). ونريد من القارئ أن يضع في الاعتبار أن CM هو عدد وفيات الأطفال تحت خمس سنوات لكل 1000 مولود حي وPGNP هو الـ GNP لكل فرد في 1980، والـ FLR مقاس كنسبة مئوية. عينتنا تحتوي على 64 بلدًا.

باستخدام الحزمة الإحصائية Eviews 3، حصلنا على النتائج التالية:

$$\begin{aligned} \widehat{\text{CM}}_i &= 263.6416 \; - \; 0.0056\ \text{PGNP}_i - \; 2.2316\ \text{FLR}_i \\ \text{se} &= (11.5932) \quad (0.0019) \qquad\qquad (0.2099) \qquad R^2 = 0.7077 \qquad (2.6.7) \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \bar{R}^2 = 0.6981^* \end{aligned}$$

حيث إن الأرقام داخل الأقواس هي تقديرات الأخطاء القياسية، وقبل البدء في تحليل نتائج هذا الانحدار، لاحظ أن معامل الميل الجزئي لـ PGNP يساوي 0.0056−.

(*) لمزيد من التفاصيل، انظر الفقرة 8.7.

ألا تلاحظ أنها نفس القيمة التي حصلنا عليها من العملية ثلاثية الخطوات التي ناقشناها في الفقرة السابقة [انظر المعادلة (5.3.7)]؟ ألا تفاجئك هذه النتيجة؟ ليس ذلك فقط ولكن الخطأين القياسيين هما تقريبًا متساويان، وقد حصلنا على ذلك بدون العملية ذات الثلاث خطوات السابق شرحها.

دعنا الآن نفسر معاملات الانحدار: 0.0056− هو معامل الانحدار الجزيئي الخاص بـ PGNP والذي يعني أنه كلما زاد الـ PGNP، مثلاً بدولار واحد فإنه في المتوسط تقل وفيات الأطفال بـ 0.0056 وحدة. وذلك مع افتراض ثبات تأثير FLR. ولإعطاء تفسير اقتصادي أكثر دقة، دعنا نقول إنه إذا زاد الـ GNP بألف دولار، فإنه في المتوسط سيقل عدد وفيات الأطفال تحت الخمس سنوات بحوالي (6.5) لكل ألف مولود حي.

المعامل 2.2316− يعني أنه في المتوسط يقل عدد وفيات الأطفال تحت الخمس سنوات بحوالي 2.23 لكل ألف مولود حي كلما زاد معدل تعلم القراءة والكتابة للمرأة بنقطة واحدة مئوية بافتراض ثبات تأثير الـ PGNP.

أما التفسير الخاص بالثابت 263، فإنه نظريًا يعني إذا كانت رقم PGNP ومعدل FLR يساوي الصفر، فإن متوسط وفيات الأطفال سيكون حوالي263 وفاة لكل ألف مولود حي. وبالطبع لابد دائمًا من إضافة نوع من الواقعية عند تفسير معنى الجزء الثابت من معادلة الانحدار. فما تستطيع فهمه فعليًا في هذا المثال، أنه إذا كانت المتغيرات المنحدرة تساوي الصفر، فإن وفيات الأطفال ستكون مرتفعة نوعًا ما، وذلك يعتبر أمرًا متوقعًا. قيمة R^2 المساوية لـ 0.71 تعني أن 71% من التباين في وفيات الأطفال يمكن تفسيره من خلال الـ PGNP والـ FLR، وهذه النسبة تعتبر نسبة عالية، حيث إن أقصى قيمة ممكنة لـ R^2 هي 1. مما يعني أن نتائج الانحدار لها قيمة ومعنى فعلي.

ماذا عن المعنوية الإحصائية للمعاملات المقدرة؟ سنتناول هذه النقطة بالتفصيل في الفصل (8). وكما سنرى، سيكون هذا الفصل امتدادًا للمواضيع التي نوقشت في الفصل (5)، والخاصة بالنماذج ثنائية المتغيرات. وكما سنرى أيضًا، هناك بعض الفروق المهمة في الاستدلال الإحصائي (أي اختبارات الفروض) بين نماذج الانحدار ثنائية المتغيرات ومتعددة المتغيرات.

الانحدار باستخدام المتغيرات القياسية : Regression on Standardized Variables

في الفصل السابق، تناولنا موضوع الانحدار باستخدام متغيرات قياسية، وقدمنا امتداداً للموضوع في حالة الانحدار متعدد المتغيرات. تذكر أن المتغير يقال إنه متغير قياسي أو متغير مأخوذ بوحدات الانحراف المعياري إذا تم التعبير عنه في صورة انحرافات عن وسطه الحسابي ومقسوم على انحرافه المعياري.

في مثالنا الخاص بوفيات الأطفال، كانت النتائج كالتالي:

$$\widehat{\text{CM}}^* = -0.2026\ \text{PGNP}_i^* - 0.7639\ \text{FLR}_i^*$$
$$\text{se} = \quad (0.0713) \qquad (0.0713) \qquad r^2 = 0.7077 \qquad (3.6.7)$$

لاحظ أن: المتغيرات (*) هي متغيرات قياسية. ولاحظ أيضاً أنه لا يوجد جزء ثابت في هذا الانحدار لأسباب سبق وتم شرحها في الفصل السابق.

كما ترى من هذا الانحدار، بافتراض ثبات FLR، فإن زيادة الانحراف المعياري في PGNP تؤدي في المتوسط لانخفاض الانحراف المعياري بـ 0.2026 في الـ CM وبالمثل بافتراض ثبات PGNP، فإن زيادة الانحراف المعياري في FLR في المتوسط تؤدي إلى انخفاض الانحراف المعياري بـ 0.7639 في الـ CM. وبشكل نسبي، فإن معدل تعلم القراءة والكتابة للمرأة له أثر أكبر على وفيات الأطفال من أثر الـ PGNP، فباستخدام الشكل القياسي للمتغيرات نجعل كل المتغيرات ذات وحدة قياس واحدة حيث إنها جميعاً لها متوسط يساوي الصفر وتباين يساوي الوحدة.

7.7 الانحدار البسيط في إطار الانحدار المتعدد.. مقدمة لتحيز التوصيف: SIMPLE REGRESSION IN THE CONTEXT OF MULTIPLE REGRESSION: INTRODUCTION TO SPECIFICATION BIAS

تذكر أن فرض (6.1.7) الخاص بنموذج الانحدار الخطي التقليدي يعني أن نموذج الانحدار المستخدم في التحليل تم توصيفه بشكل «صحيح»، أي أنه لا يوجد تحيز في التوصيف أو خطأ في التوصيف (انظر الفصل (3)، حيث توجد بعض الملاحظات المبدئية في ذلك الموضوع). فعلى الرغم من أن موضوع خطأ التوصيف ستتم مناقشته بشكل كامل في الفصل (13)، إلا أن المثال التوضيحي المعطى في الفقرة السابقة يعتبر فرصة جيدة لاستعراض أهمية هذا الغرض (6.1.7) ليس ذلك فقط وإنما أيضاً فرصة جيدة لإلقاء الضوء على معنى معاملات الانحدار الجزيئية وعمل مقدمة بسيطة لموضوع تحيز التوصيف.

بافتراض أن (1.6.7) هو النموذج «الصحيح» الذي يصف وفيات الأطفال وعلاقتها بالـ PGNP ومعدل تعلم القراءة والكتابة للمرأة (PLR). والآن افترض أننا سنتجاهل FLR ونقدر الانحدار البسيط التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \tag{1.7.7}$$

حيث $CM = Y$ و $PGNP = X_2$.

بما أن (1.6.7) هو النموذج الصحيح، فإن تقدير (1.7.7) سيكون قيمة خطأ في التوصيف، الخطأ الخاص بحذف المتغير X_3، معدل تعلم القراءة والكتابة للمرأة.

لاحظ أننا نستخدم رموزاً مختلفة للمعاملات (α's) في (1.7.7) للتفريق بينها وبين المعاملات الصحيحة (β's) المعطاة في (1.6.7).

والآن هل ستعطي α_2 مقدراً غير متحيز للتأثير الحقيقي لـ PGNP والذي حصلنا عليه من قبل من خلال β_2 في النموذج (1.6.7)؟ بمعنى آخر هل $E(\hat{\alpha}_2) = \beta_2$، حيث إن $\hat{\alpha}_2$ هي القيمة المقدرة لـ α_2؟

بمعنى آخر، هل معامل PGNP في (1.7.7) سيعطي مقدراً غير متحيز للتأثير الحقيقي لـ PGNP على CM، مع الوضع في الاعتبار أننا حذفنا المتغير X_3 (FLR) من النموذج؟ كما نتوقع، في العموم فإن $\hat{\alpha}_2$ لن يكون مقدراً غير متحيز للقيمة الحقيقية β_2. ولإعطاء لمحة عن التحيز، دعنا نقوم بإجراء انحدار (1.7.7)، والذي يعطي النتائج التالية:

$$\begin{aligned} \widehat{CM}_i &= 157.4244 \quad - \quad 0.0114\ PGNP_i \\ se &= \ (9.8455) \qquad (0.0032) \qquad\qquad r^2 = 0.1662 \end{aligned} \tag{2.7.7}$$

لاحظ الآن الملاحظات التالية، والخاصة بهذا الانحدار، مقارنةً بالانحدار المتعدد «الصحيح» المعطى في (1.6.7):

1 - باستخدام القيم المطلقة (بمعنى استبعاد الإشارة) فإن معامل PGNP زاد من 0.0056 إلى 0.0114، الضعف تقريباً.

2 - الأخطاء القياسية اختلفت.

3 - قيم الجزء الثابت من الانحدار اختلفت.

4 - قيم r^2 اختلفت بشكل كبير، وعادة ما يكون ذلك صحيحاً، حيث إنه كلما زاد عدد المتغيرات المنحدرة كلما زادت قيمة r^2.

والآن دعنا نقترض أننا سنقوم بعمل انحدار لوفيات الأطفال على معدل حرية المرأة مع تجاهل أو استبعاد أثر الـ PGNP، سنحصل على النتائج التالية:

$$\widehat{\text{CM}}_i = 263.8635 - 2.3905\ \text{FLR}_i$$
$$\text{se} = (21.2249) \quad (0.2133) \qquad r^2 = 0.6696 \qquad (3.7.7)$$

الآن مرة أخرى إذا قارنت نتائج هذا الانحدار (غير الموصف بشكل صحيح) مع نموذج الانحدار المتعدد (الصحيح)، ستجد أن النتائج مختلفة، على الرغم من أن الاختلاف هنا ليس كبيراً مثل حالة انحدار (2.7.7).

النقطة الأجدر بالملاحظة هنا هي العواقب الخطيرة الممكن حدوثها في حالة التوفيق الخاطئ للنموذج. سنتناول هذه النقطة بتفصيل أكثر في الفصل (13)، والخاصة بأخطاء التوصيف.

7.8 R^2 و R^2 المعدلة: R^2 AND THE ADJUSTED R^2

من الخصائص المهمة لـ R^2 أنه دالة غير تناقصية في عدد المتغيرات المفسرة أو المنحدرة الموجودة في النموذج، أي أنه كلما زاد عدد المتغيرات المنحدرة كلما زادت قيمة R^2 ولا تقل أبداً. وبعبارة أخرى فإضافة متغير X جديد لن تقلل من R^2.

فعلى سبيل المثال، قارن انحدار (2.7.7) أو (3.7.7) مع (2.6.7) حتى يسهل فهم المقصود من هذه الفقرة دعنا نتذكر تعريف معامل التحديد:

$$\begin{aligned} R^2 &= \frac{\text{ESS}}{\text{TSS}} \\ &= 1 - \frac{\text{RSS}}{\text{TSS}} \qquad (1.8.7) \\ &= 1 - \frac{\sum \hat{u}_i^2}{\sum y_i^2} \end{aligned}$$

الآن $\sum y_i^2$ مستقل عن عدد المتغيرات المفسرة X في النموذج، حيث إنه ببساطة يساوي $\Sigma(Y_i - \bar{Y})^2$. أما RSS، $\sum \hat{u}_i^2$ فإنه يعتمد على عدد المتغيرات المنحدرة الموجودة في النموذج. وبالتالي مبدئيًا يتضح أنه كلما زاد عدد المتغيرات المفسرة X، فإنه غالبًا ما سيقل $\sum \hat{u}_i^2$ (على الأقل لن يزيد)، وبما أن R^2 معرفة كما في (1.8.7) لن تزيد. وبالتالي اعتماداً على ذلك، فإنه في مقارنة نموذجي انحدار لهما نفس

المتغير التابع ولكن مختلفين في عدد المتغيرات المفسرة X، فعلى الفرد أن يختار النموذج الذي له أعلى R^2.

لمقارنة قيمتين للـ R^2، لابد أن يؤخذ في الاعتبار عدد متغيرات الـ X الموجودة في النموذج. وذلك يمكن بسهولة إذا اعتبرنا معامل التحديد البديل التالي:

$$\bar{R}^2 = 1 - \frac{\sum \hat{u}_i^2 / (n-k)}{\sum y_i^2 / (n-1)} \quad (2.8.7)$$

حيث k = عدد المعلمات الموجودة في النموذج بإضافة إلى الجزء الثابت (في انحدار ثلاثي المتغيرات يكون $k = 3$، لماذا؟) فإن $\bar{R}^2$ المعرفة سابقًا تعرف باسم R^2 المصححة، ويرمز لها $\bar{R}^2$. معنى كلمة المصححة أي أنها مصححة وفقًا لدرجات الحرية الخاصة بمجموع المربعات الداخل في المعادلة (1.8.7): $\sum \hat{u}_i^2$ لها $(n-k)$ درجات حرية في نموذج يحتوي على k معلمة، وجزء ثابت، و $\sum y_i^2$ له $(n-1)$ درجة حرية (لماذا؟). بالنسبة لحالة وجود ثلاثة متغيرات، نعرف أن $\sum \hat{u}_i^2$ له $(n-3)$ درجة حرية.

المعادلة (2.8.7) يمكن كتابتها كالتالي:

$$\bar{R}^2 = 1 - \frac{\hat{\sigma}^2}{S_Y^2} \quad (3.8.7)$$

حيث $\hat{\sigma}^2$ هي تباين البواقي، وهي مقدر غير متحيز للقيمة الحقيقية σ^2 و S_y^2 هو تباين العينة لـ Y.

من السهل إثبات وجود علاقة بين $\bar{R}^2$ و R^2 حيث بالتعويض عن (1.8.7) في (2.8.7) نحصل على

$$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k} \quad (4.8.7)$$

ويظهر بوضوح من المعادلة (4.8.7) أن (1) لكل $k > 1$ فإن $\bar{R}^2 > R^2$ أي أنه مع زيادة عدد المتغيرات المفسرة X فإن R^2 المصححة تزيد بدرجة أقل من R^2 غير المصححة و (2) $\bar{R}^2$ ممكن أن تكون سالبة على الرغم من أن R^2 هي بالضرورة قيمة غير سالبة[10]. في حالة وجود $\bar{R}^2$ سالبة في أحد التطبيقات العملية يتم اعتبارها تساوي الصفر.

(10) لاحظ عمومًا أنه إذا كان $R^2=1$ فإن $\bar{R}^2 = R^2 = 1$ وعندما تكون $R^2 = 0$ فإن $\bar{R}^2=(1-k)/(n-k)$ والتي تكون سالبة في هذه الحالة إذا كان $k > 1$.

أي قيمة لـ R^2 يجب استخدامها عمليًا؟ كما قال Theil فإن:

من الأفضل عمليًا استخدام $\bar{R}^2$ بدلاً من R^2 حيث R^2 تعطي صورة أفضل عن جودة توفيق النموذج عما هي في الواقع خصوصًا عندما يكون عدد المتغيرات المفسرة ليس قليلاً جدًا مقارنةً مع عدد المشاهدات[11].

ولكن وجهة نظر Theil لا تلقى موافقة عامة، حيث إنه لم يقدم دليلاً نظريًا على «تفوق» $\bar{R}^2$. فعلى سبيل المثال، فإن Goldberger اقترح قيمة R^2 التالية، وأطلق عليها اسم R^2 المعدلة، وقال إن لها مميزات هي الأخرى كالتالي[12]:

$$\text{Modified } R^2 = (1 - k/n)R^2 \tag{5.8.7}$$

ونصح باستخدام أيضًا R^2، وتحديد الـ n والـ k وترك القارئ يختار كيفية تصحيح الـ R^2 إما من خلال $\bar{R}^2$ أو الـ R^2 المعدلة.

وبغض النظر عن ذلك، فإن R^2 المصححة كما في (4.8.7) تعطي دائمًا في نتائج معظم الحزم الإحصائية بالإضافة إلى R^2 التقليدية. وننصح القارئ باستخدام $\bar{R}^2$ كإحصاء آخر يساعد في الحكم على النموذج.

وبالنسبة لنموذج الانحدار الخاص بوفيات الأطفال (2.6.7)، على القارئ أن يثبت أن $\bar{R}^2$ تساوي 0.6981، مع الوضع في الاعتبار أن في هذا المثال $(n-1) = 63$ و $(n-k) = 60$. كما هو متوقع فإن $\bar{R}^2$ المساوية لـ 0.6981 أقل من R^2 المساوية لـ 0.7077 بالإضافة إلى R^2 و R^2 المصححة كمقياس لجودة توفيق النموذج، هناك طرق أخرى تستخدم عادة للحكم على مدى دقة نموذج الانحدار. وتعتبر طريقة معلومات Akaike وطريقة التنبؤ لـ Amemiya طريقتين أخريين للحكم بين النماذج المتنافسة.

وسنتناول هاتين الطريقتين بالتفصيل عند التعرض لمشكلة كيفية اختيار النموذج في فصل لاحق (انظر الفصل 13).

---

(11) Henri Theil, Introduction to econometrics, Prentice Hall, Englewood Cliffs, N. J., 1978, p. 135.

(12) لدراسة أكثر عمقًا عن R^2، انظر

S. Gameron, "Why is the R Squared Adjusted Reported?", Joumal of Quantitative Economics, vol. 9. no. 1, January 1993, pp. 183-186.

وقال في هذا البحث أن "R^2 ليست اختباراً إحصائيًا ولا يوجد تفسير واضح لها لاستخدامها كإحصاء وصفي.. وأخيراً لابد من وضوح فكرة عدم فعالية R^2 كأداة للحكم على دقة النموذج" (صفحة 186).

المقارنة بين قيمتين لـ R^2 : Comparing two R^2 Values

من الضروري ملاحظة أنه عند مقارنة نموذجين على أساس معاملات التحديد سواء معامل تحديد مصحح أم لا، لابد أن يكون حجم العينة n والمتغير التابع متساويين في النموذجين، أما المتغيرات المفسرة فممكن أن تكون مختلفة. وبالتالي بالنسبة للنماذج التالية:

$$\ln Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (6.8.7)$$

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \quad (7.8.7)$$

لا يمكن مقارنة R^2 للنموذجين. السبب هو التالي: بالتعريف، R^2 تقيس نسبة التباين في المتغير التابع التي يمكن تفسيرها من خلال المتغيرات المفسرة. وبالتالي في (6.8.7) R^2 تقيس نسبة التباين في $\ln Y$ المفسرة من خلال X_2 و X_3، في حين أن في (7.8.7) تقيس R^2 نسبة التباين في Y، وبالتالي الشيئان لا يتساويان: كما لاحظنا في الفصل (6)، التغير في $\ln Y$ يعطي تغيراً نسبيًا في Y، في حين التغير في Y يعطي تغيراً مطلقًا.

وبالتالي، فإن $\text{var}\ \hat{Y}_i\ /\text{var}\ Y_i$ لا يساوي $\text{var}\ (\widehat{\ln Y_i})/\text{var}(\ln Y_i)$ وبالتالي فمعاملات التحديد غير متساوية.[13]

كيف يمكن إذن مقارنة R^2 الخاصة بهذين النموذجين إذا لم يكونا على نفس الشكل؟ للإجابة عن هذا السؤال دعنا نستعرض أولاً المثال الرقمي التالي.

---

(13) من تعريف R^2، نعرف أن

$$1 - R^2 = \frac{\text{RSS}}{\text{TSS}} = \frac{\sum \hat{u}_i^2}{\sum (Y_i - \bar{Y})^2}$$

بالنسبة للنماذج الخطية أما

$$1 - R^2 = \frac{\sum \hat{u}_i^2}{\sum (\ln Y_i - \overline{\ln Y})^2}$$

بالنسبة للنموذج log. وبما أن المقام في الجانب الأيمن من المعادلة في المعادلتين السابقتين غير متساو لا تستطيع مقارنة R^2 للنموذجين معًا.

كما يتضح من مثال 2.7، للنموذج الخطي، فإن RSS=0.1491 (مجموع مربعات البواقي لاستهلاك القهوة) وبالنسبة للنموذج اللوغاريتمي log، فإن RSS= 0.0226 (مجموع مربعات بواقي لوغاريتم استهلاك القهوة). هذه البواقي مختلفة في درجتها ولا يمكن مقارنتها معًا.

مثال 2.7

استهلاك القهوة في الولايات المتحدة، 1970–1980

اعتبر البيانات المعطاة في جدول (1.7). البيانات خاصة باستهلاك أكواب من القهوة في اليوم الواحد (Y) وسعر القهوة (X) في الولايات المتحدة في الفترة من 1970-1980 باستخدام الـ OLS حصلنا على نتائج الانحدار التالية:

$$\hat{Y}_t = 2.6911 - 0.4795X_t$$
$$\text{se} = (0.1216) \quad (0.1140) \qquad \text{RSS} = 0.1491;\ r^2 = 0.6628 \qquad (8.8.7)$$

هذه النتائج لها معنى اقتصادي، فكلما زاد سعر القهوة في المتوسط يقل استهلاك القهوة بحوالي نصف فنجان في اليوم الواحد. قيمة r^2 المساوية لـ 0.66 تعني أن سعر القهوة يفسر حوالي 66% من التباين في استهلاك القهوة.

ويمكن للقارئ أن يثبت أن معامل الميل له معنوية إحصائية.

من نفس البيانات، تم تقدير نموذج ثنائي اللوغاريتم أو ذو المرونة الثانية التالي:

$$\widehat{\ln Y_t} = 0.7774 - 0.2530 \ln X_t$$
$$\text{se} = (0.0152) \quad (0.0494) \qquad \text{RSS} = 0.0226;\ r^2 = 0.7448 \qquad (9.8.7)$$

وبما أن هذا النموذج هو ثنائي اللوغاريتم، فإن معامل الميل يعطي تقديراً مباشراً لمعامل مرونة السعر. وبالتالي في مثالنا الحالي، فإن هذا المعامل يعني أنه كلما زاد سعر القهوة بـ 1% فإن في المتوسط استهلاك القهوة يقل بحوالي 0.25% في اليوم الواحد. تذكر أنه في النموذج الخطي (7.8.8) معامل الميل يعطي فقط معدل التغير في استهلاك القهوة بالنسبة للسعر. (كيف يمكن تقدير مرونة السعر بالنسبة للنموذج الخطي؟).

جدول (1.7) – استهلاك القهوة في الولايات المتحدة (Y) وعلاقته بمتوسط سعر القهوة (X) (*)، 1970-1980

Table - V. S. Coffee Comsumption (Y) in velation to average real netail price (X), 1970–1980(*)

Year	Y, Cups per person per day	X, $ per lb
1970	2.57	0.77
1971	2.50	0.74
1972	2.35	0.72
1973	2.30	0.73
1974	2.25	0.76
1975	2.20	0.75
1976	2.11	1.08
1977	1.94	1.81
1978	1.97	1.39
1979	2.06	1.20
1980	2.02	1.17

* لاحظ أن: السعر الاسمى تم قسمته على مؤشر سعر المستهلك (CPI) للغذاء والمشروبات، 100=1967.

المصدر: بيانات (Y) تم الحصول عليها من: National loffee dninking study, Data group El kins part, penn., 1981.

وبيانات X الاسمية (أي X في الأسعار الحالية) تم الحصول عليها من: Nielsen Food index, A. C. Nielsen, New york, 1981

أشكر Slott E. Sandberg على تجميع البيانات.

قيمة r^2 حوالي 0.74 وهي تعني أن حوالي 74% من التباين في لوغاريتم الطلب على القهوة يمكن تفسيره من خلال التباين في لوغاريتم سعر القهوة.

وبما أن قيمة r^2 في النموذج اللوغاريتمي الخطي، فسيتم اختيار النموذج الأخير الذي له r^2 أعلى. ولكن للأسباب السابق شرحها، لا يمكننا فعل ذلك، ولكن إذا أردت مقارنة قيمتين r^2، فيمكنك عمل التالي:

1 - احصل على $\widehat{\ln Y}$ من (9.8.7) لكل مشاهدة، أي احصل على القيمة المقدرة للوغاريتم كل مشاهدة في النموذج. ثم احصل على معكوس اللوغاريتم antilog لهذه القيم واحسب r^2 باستخدام هذه القيم antilog وقيم Y_t الفعلية بنفس الطريقة المستخدمة في المعادلة (15.5.3) ∴ هذه القيمة لـ r^2 يمكن مقارنتها مع قيمة r^2 التي حصلت عليها من النموذج الخطي (8.8.7).

2 - وكأسلوب بديل، يمكنك افتراض أن كل قيم Y هي قيم موجبة ثم احصل على لوغاريتم هذه القيمة، $\ln Y$ واحصل على القيم المقدرة $\hat{Y}_t$ من النموذج الخطي واحسب لوغاريتم هذه القيم المقدرة (أي $\ln \hat{Y}_t$) واحسب بعد ذلك r^2 بين ($\ln Y_t$ و $\ln \hat{Y}_t$) بنفس الطريقة المستخدمة في المعادلة (14.5.3). هذه القيمة لـ r^2 يمكن مقارنتها مع r^2 التي حصلت عليها من (9.8.7).

بالنسبة لمثالنا الخاص بالقهوة، فقد عرضنا البيانات الخام الضرورية لحساب r's ببحث يمكن مقارنتهما بشكل صحيح في الجدول (2.7). لمقارنة قيمة r^2 الخاصة بالنموذج الخطي (8.8.7) مع نظريتها الخاصة بـ (9.8.7)، سنقوم أولاً بالحصول على لوغاريتم ($\hat{Y}_t$) [المعطى في العمود (6) من الجدول (2.7)] ثم نحصل على لوغاريتم القيمة الحقيقية للـ Y [المعطى في العمود (5) من نفس الجدول] ثم نقارن هاتين المجموعتين من القيم باستخدام المعادلة (14.5.3) النتيجة هي قيمة r^2 المساوية لـ 0.7318 والتي يمكن مقارنتها الآن مع قيمة r^2 للنموذج اللوغاريتمي الخطي المساوية لـ 0.7318 والتي يمكن مقارنتها الآن مع قيمة r^2 للنموذج اللوغاريتمي الخطي المساوية لـ 0.7448. ونرى الآن أن الفرق بين القيمتين صغيرة جداً.

على الجانب الآخر، إذا أردنا مقارنة قيمة r^2 الخاصة بالنموذج اللوغاريتمي الخطي مع النموذج الخطي، نحصل على $\widehat{\ln Y_t}$ لكل مفردة من (9.8.7) [المعطى في العمود (3) من الجدول] ونحصل على قيم اللوغاريتم العكسية [المعطاة في العمود (4) من الجدول] وأخيراً نسب r^2 من القيمة العكسية للوغاريتم وقيم Y الفعلية باستخدام المعادلة (14.5.3). هذا سيعطي r^2 مساوية لـ 0.7187 والتي تعتبر أكبر قليلاً مما حصلنا عليه من النموذج الخطي (8.8.7) والمساوية لـ 0.6628 باستخدام أي من الطريقتين، يبدو أن النموذج اللوغاريتمي الخطي يعطي نتائج أفضل نوعًا ما.

جدول (2.7) البيانات الخام الخاصة بمقارنة قيمتين الـ R^2

Raw data for comparing two R^2 values

Year	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	Antilog of $\widehat{\ln Y_t}$ (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

لاحظ أن: العمود (1): قيم Y الفعلية من جدول (1.7)

العمود (2): قيم Y المقدرة من النموذج الخطي (8.8.7)

العمود (3): قيم لوغاريتم الـ Y المقدرة من النموذج اللوغاريتمي الثنائي (9.8.7)

العمود (4): قيم اللوغاريتم العكسية للعمود (3).

العمود (5): لوغاريتم الـ Y الموجودة في العمود (1).

العمود (6): لوغاريتم قيم الـ $\hat{Y}_t$ الموجودة في العمود (2).

تعيين قيمة R^2 بين المتغيرات المنحدرة : Allocating R^2 Among Regressors

دعنا نعود إلى مثالنا الخاص بوفيات الأطفال. رأينا في (2.6.7) أن المتغيرين المنحدرين PGNP و FLR يفسران 0.7077 أو 70.77% من التباين في وفيات الأطفال، ولكن الآن دعنا نعتبر الانحدار (2.7.7) حيث أسقطنا متغير FLR، وكنتيجة لذلك انخفضت قيمة r^2 إلى 0.1662. هل هذا يعني أن الفرق بين قيمتين r^2 المساوي لـ 0.5415 (0.7077–0.1662) يعبر عن المتغير الذي تم إهماله FLR؟ على الجانب الآخر، إذا اعتبرنا الانحدار (3.7.7) حيث أسقطنا المتغير PGNP قيمة r^2 انخفضت إلى 0.6696. هل هذا يعني أن الفرق في قيمة r^2 المساوي لـ 0.0381 (0.7077–0.6696) يعود إلى المتغير المحذوف PGNP؟

السؤال إذن هو: هل يمكن تعيين قيمة R^2 المساوية لـ 0.7077 بين المتغيرين المنحدرين، PGNP و FLR بالطريقة السابق ذكرها؟ للأسف لا نستطيع فعل ذلك، حيث إن ذلك يعتمد على ترتيب وجود هذه المتغيرات كما سبق وشرحنا. جزء من هذه المشكلة هنا يعتمد على أن المتغيرين مرتبطان، فمعامل الارتباط بينهما يساوي 0.2685 (أثبت ذلك باستخدام البيانات المعطاة في جدول 4.6). في معظم المواقف

التطبيقية المرتبطة بعدد من المتغيرات المنحدرة، فيمثل الارتباط بين هذه المتغيرات مشكلة حقيقية بالطبع ستكون المشكلة أكثر خطورة إذا كان هناك ارتباط خطي تام بين المتغيرات المنحدرة.

والنصيحة العملية المثلى هي وجوب الحذر والدقة عند تعيين قيمة R^2 بين عدد من المتغيرات المنحدرة المكونة لها.

«لعبة» تعظيم قيمة $\bar{R}^2$: $\bar{R}^2$ The "Game" of Maxmizing

في ختام هذه الفقرة، لابد من توضيح التحذير التالي: أحيانًا يقوم الباحثون بعملية تعظيم أو تكبير قيمة $\bar{R}^2$، وبالتالي اختيار النموذج الذي له أعلى $\bar{R}^2$. ولكن ذلك قد يكون خطيرًا حيث إنه في تحليل الانحدار هدفنا ليس الحصول على $\bar{R}^2$ مرتفعة ولكن الحصول على تقديرات لمعاملات الانحدار الخاصة بالمجتمع وعمل استدلال إحصائي لهذه المعلمات. في التحليل التطبيقي ليس مستغربًا أن تحصل على $\bar{R}^2$ مرتفعة جدًا ولكن نجد بعض معاملات الانحدار إما غير معنوية إحصائيًا أو لها إشارة تختلف عن الإشارة المتوقعة لمثل هذا المعامل. وبالتالي، على القارئ أن يهتم أكثر بالمعنى النظري أو المنطقي الخاص بالمتغيرات المفسرة وعلاقتها بالمتغير التابع ومدى معنوية هذه المتغيرات. إذا حصلنا عند ذلك على $\bar{R}^2$ مرتفعة، فإن هذا يعتبر مؤشرًا جيدًا أما العكس فليس صحيحًا بمعنى أنه إذا كانت $\bar{R}^2$ منخفضة فذلك لا يعني بالضرورة أن النموذج سيء.[14]

وفي حقيقة الأمر، فإن Goldberger حذر من الدور الذي تلعبه $\bar{R}^2$ قائلاً:

من وجهة نظري، فإن $\bar{R}^2$ يجب أن يكون لها دور بسيط في تحليل الانحدار كمقياس لجودة توفيق تقديرات العينة الخاصة بـ LS (المقدرات العنصري).

---

(14) بعض الكتاب يؤكدون على استخدام R^2 كمقياس لجودة التوفيق واستخدامها أيضًا كمقارنة بين قيمتين أو أكثر للـ R^2 انظر

Christopher H. Achen, Interpreting and Using Regression, Sage Publications, Beverly Hills, Calif., 1982, pp. 58-67, and C. Granger and P. Newbold, "R^2 and the Transformation of Regression Variables," Journal of Econometrics, vol. 4, 1976, pp. 205-210.

وبهذه المناسبة استخدام أعلى R^2 كأسلوب لاختيار أفضل نموذج قدم ما يعرف باسم Pretest bias (التحيز المسبق)، والذي قد يؤثر سلبيًا بشكل كبير على مقدرات الـ OLS الخاصة بنموذج الانحدار الخطي التقليدي. لمزيد من التفاصيل عن هذه النقطة يمكن للقارئ أن يستعين بالتالي:

George G. Judge, Carter R. Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, Introduction to the Theory and Practice of Econometrics, John Wiley, New York, 1982, Chap. 21.

لا يوجد في نماذج الـ [CLRM] ما يتطلب ضرورة وجود $\bar{R}^2$ مرتفعة. وبالتالي فوجود قيمة مرتفعة لـ$\bar{R}^2$ ليس دليلاً لصالح النموذج والقيمة المنخفضة لـ$\bar{R}^2$ ليست دليلاً ضد النموذج.

في الواقع، الشيء الأكثر أهمية الخاص بـ R^2 هي أنها غير مهمة في نماذج الـ CR، فنموذج الـCR يهتم بمعالم المجتمع وليس جودة توفيق العينة. وإذا كان الفرد مصرًا على استخدام مقياس لنجاح التنبؤ (أو فشله)، فبالتالي σ^2 قد تكون كافية: مفهومًا فإن المعلمة σ^2 هي مربع توقع خطأ التنبؤ الذي سيحدث إذا استخدمنا الـ CEF الخاص بالمجتمع (PRE) كمتغير يستخدم للتنبؤ. بعبارة أخرى، فإن مربعات الأخطاء القياسية للتنبؤ باستخدام قيم معينة للـ x (المتغيرات المنحدرة) قد تكون مقيدة للحكم على جودة توفيق النموذج.[15]

9.7 مثال 3.7: دالة إنتاج COBB - DOUGLAS: المزيد عن شكل الدالة: EXAMPLE 7.3: THE COBB-DOUGLAS PRODUCTION FUNCTION: MORE ON FUNCTIONAL FROM

في الفقرة 4.6 أوضحنا كيف يمكن باستخدام تحويلات مناسبة تحويل العلاقة غير الخطية إلى خطية، وبالتالي تستطيع العمل في إطار نموذج الانحدار الخطي البسيط. التحويلات المتعددة التي تمت مناقشتها من قبل والخاصة بحالة وجود متغيرين اثنين فقط يمكن أن تمد لتشمل نماذج الانحدار المتعدد. وسنقدم في هذه الفقرة تحويلات خاصة بوجود أكثر من متغيرين في نماذج اللوغاريتم الخطية log-linear Models بعض التحويلات الأخرى موجودة في التمارين، وفي الأمثلة التوضيحية التي سيتم مناقشتها في باقي أجزاء الكتاب. المثال الذي سنتناوله هو دالة إنتاج Cobb-Douglas في شكلها العشوائي يمكن كتابتها:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \qquad (1.9.7)$$

حيث Y = الناتج

X_2 = مدخل العمالة

X_3 = مدخل رأس المال

u = مقدار الخطأ العشوائي

e = أساس اللوغاريتم الطبيعي

(15) Arther S. Goldberger, op. cit., pp. 177-178.

من المعادلة (1.9.7) واضح أن العلاقة بين الناتج والمدخلات غير خطية. عمومًا إذا استخدمنا التحويلة اللوغاريتم على هذا النموذج نحصل على:

$$\begin{aligned} \ln Y_i &= \ln \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \\ &= \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \end{aligned} \qquad (2.9.7)$$

حيث $\beta_0 = \ln \beta_1$

والآن هذا النموذج خطي في المعلمات β_0، β_2، β_3، وبالتالي هو نموذج انحدار خطي. لاحظ أنه على الرغم من أنه نموذج غير خطي في المتغيرين Y و X إلا أنه خطي في لوغاريتم هذه المتغيرات. باختصار فإن (2.9.7) هي معادل اللوغاريتم – اللوغاريتم، اللوغاريتم المزدوج أو اللوغاريتم الخطي لنموذج اللوغاريتم الخطي ثنائي المتغيرات (3.5.6) ولكن في حالة وجود أكثر من متغيرين.

خصائص دالة إنتاج Cobb-Douglas المشهورة هي كالتالي:

1 – β_2 هي المرونة الجزيئية للناتج بالنسبة لمدخل العمالة، بمعنى أنه يقيس نسبة التغير في الناتج لكل 1%، مثلاً، تغير في مدخل العمالة، بافتراض ثبات رأس المال (انظر تمرين 9.7).

2 – وبالمثل، فإن β_3 هو المرونة الجزيئية للناتج با لنسبة لمدخل رأس المال بافتراض ثبات العمالة

3 – مجموع $(\beta_2 + \beta_3)$ يعطي معلومات عن مدى استجابة الناتج لتغير نسبي في المدخلات، إذا كان المجموع يساوي 1، فإن ذلك يعني أنه بمضاعفة المدخلات ستتم مضاعفة الناتج، وثلاثة أمثال المدخلات سيعطي ثلاثة أمثال الناتج وهكذا. إذا كان المجموع أقل من 1 فيكون المعدل متناقصًا، بمعنى أن مضاعفة المدخلات سيكون أقل من مضاعفة الناتج. وأخيرًا إذا كان المجموع أكبر من 1 فسيكون المعدل متزايدًا بمعنى أن مضاعفة المدخلات سيكون أكثر من مضاعفة الناتج.

وقبل استعراض المزيد من التفاصيل، لاحظ أنه عندما يكون لديك نموذج انحدار لوغاريتمي خطي يحتوي على أي عدد من المتغيرات، فإن معامل كل متغير من المتغيرات المفسرة X يقيس المرونة (الجزيئية) للمتغير التابع Y بالنسبة لهذا المتغير. وبالتالي إذا كان لديك k متغير في نموذج لوغاريتمي خطي كالتالي:

$$\ln Y_t = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + \ldots + \beta_k \ln X_{ki} + u_i \qquad (3.9.7)$$

فإن كلاً من معاملات الانحدار الجزيئية من β_2 حتى β_k تمثل المرونة (الجزيئية) لـ Y بالنسبة للمتغيرات X_2 حتى X_k.[16]

لشرح دالة إنتاج Cobb-Douglas، حصلنا على البيانات الموجودة في جدول (3.7)، هذه البيانات خاصة بالقطاع الزراعي في تايوان في الفترة من 1958–1972.

بافتراض أن النموذج (2.9.7) مستوفي كل شروط نموذج الانحدار الخطي التقليدية[17]. نحصل على الانحدار التالي باستخدام طريقة OLS (انظر الملحق A7، الفقرة 5.A7 للاطلاع على مخرجات الحاسب الآلي):

جدول (3.7) الناتج الإجمالي الحقيقي، أيام العمل، ومدخل رأس المال الحقيقي في قطاع الزراعة في تايوان 1972–1958

Year	Real gross product (millions of NT $)*, Y	Labor days (millions of days), X_2	Real capital input (millions of NT $), X_3
1958	16,607.7	275.5	17,803.7
1959	17,511.3	274.4	18,096.8
1960	20,171.2	269.7	18,271.8
1961	20,932.9	267.0	19,167.3
1962	20,406.0	267.8	19,647.6
1963	20,831.6	275.0	20,803.5
1964	24,806.3	283.0	22,076.6
1965	26,465.8	300.7	23,445.2
1966	27,403.0	307.5	24,939.0
1967	28,628.7	303.7	26,713.7
1968	29,904.5	304.7	29,957.8
1969	27,508.2	298.6	31,585.9
1970	29,035.5	295.5	33,474.5
1971	29,281.5	299.0	34,821.8
1972	31,535.8	288.1	41,794.3

المصدر: Thomas Pei-Fan Chen, "Economic Growth and Structural Change in Taiwan - 1952-1972, A Production Function Aproach," unpublished Ph. D. thesis, Dept. of Economics, Gradutate Center, City University of New York, June 1976, Table II.

* دولار تايوان الجديد.

(16) لرؤية ذلك قم بتفاضل (3.9.7) جزئيًا بالنسبة للوغاريتم كل متغير من الـ X. وبالتالي ستحصل على: $\partial \ln Y/\partial \ln X_2 = (\partial Y/\partial X_2)(X_2/Y) = \beta_2$ والذي يمثل بالتعريف مرونة Y بالنسبة لـ X_2 و $\partial \ln Y/\partial \ln X_3 = (\partial Y/\partial X_3)(X_3/Y) = \beta_2$ والذي يمثل مرونة Y بالنسبة لـ X_3.

(17) لاحظ أن في دالة إنتاج Cobb Douglas (11.9.7) تم تقديم مقدار الخطأ العشوائي في شكل خاص بحيث يسمح بدخوله بشكل خطي في إطار تحويله اللوغاريتم الناتجة، للمزيد عن ذلك. انظر الفقرة 9.6.

$$\widehat{\ln Y_i} = -3.3384 + 1.4988 \ln X_{2i} + 0.4899 \ln X_{3i}$$
$$(2.4495) \quad (0.5398) \quad (0.1020)$$
$$t = (-1.3629) \quad (2.7765) \quad (4.8005)$$
$$R^2 = 0.8890 \qquad \text{df} = 12$$
$$\bar{R}^2 = 0.8705 \qquad (4.9.7)$$

من المعادلة (4.9.7) نرى أن القطاع الزراعي التايواني في الفترة 1958–1972 كانت مرونة الناتج بالنسبة للعمالة ورأس المال هي 1.4988 و0.4899 بالترتيب. بمعنى آخر، على مدار فترة الدراسة وبافتراض ثبات رأس المال، فإن الزيادة بنسبة 1% في العمالة يؤدي إلى زيادة في متوسط الناتج بحوالي 1.5%. وبالمثل بافتراض ثبات العمالة، فإن الزيادة نسبة 1% في رأس المال تؤدي إلى زيادة في متوسط الناتج بحوالي 0.5%. وبإضافة المرونتين نحصل على 1.9887 وذلك يمثل معامل العائد. وكما هو مثبت، خلال فترة الدراسة، فإن قطاع الإنتاج الزراعي في تايوان كان يتميز بالزيادة في معدل العائد[18].

من وجهة النظر الإحصائية البحتة، فإن معادلة الانحدار المقدرة لها جودة توفيق عالية بالنسبة للبيانات محل الدراسة. قيمة R^2 المساوية لـ 0.8890 تعني أن حوالي 89% من التباين في (لوغاريتم) الناتج يمكن تفسيره من خلال (لوغاريتم) العمالة ورأس المال. في الفصل 8، سنرى كيف يمكن استخدام الأخطاء القياسية المقدرة لاختبار الفرض الخاص بالقيم «الحقيقية» لمعاملات دالة إنتاج Cobb Douglas بالنسبة للاقتصاد التايواني.

10.7 نماذج الانحدار المتعدد الحدود:
POLYNOMIAL REGRESSION MODELS

دعنا الآن نتطرق إلى موضوع نماذج الانحدار المتعدد، نماذج الانحدار المتعدد الحدود، حيث إنها تتواجد بكثرة في الأبحاث الاقتصادية خصوصًا المتعلقة بدوال التكلفة والإنتاج. كمقدمة لمثل هذه النماذج دعنا أولاً نوسع من مدى النماذج، حتى يمكننا استخدام نموذج الانحدار الخطى التقليدي ببساطة ويسر.

لتوضيح الأفكار، دعنا نستعرض الشكل (1.7)، والذي يربط بين التكلفة الحدية (MC) للإنتاج (Y) في المدى القصير لسلعة ما مع مستوى الناتج (X).

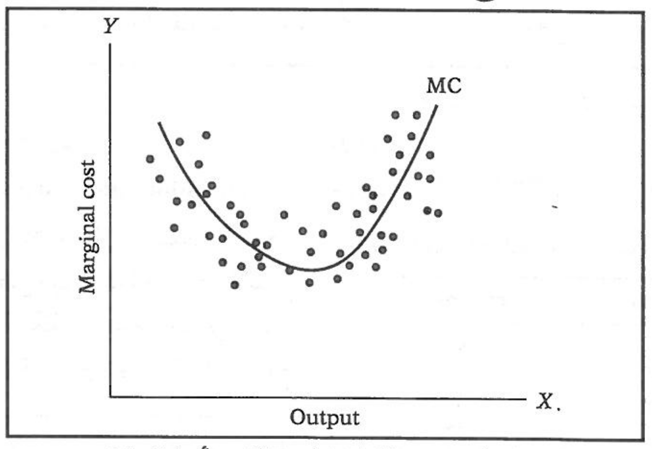
(18) نحن لم نتطرق إلى سؤال مدى تناسب النموذج مع الجانب النظري أو سؤال إمكانية قياس العائد من خلال بيانات السلاسل الزمنية.

من شكل رسمه منحنى MC، الذي يأخذ الشكل U، نرى أن العلاقة بين MC والناتج علاقة غير خطية. إذا أردنا تحديد هذه العلاقة من شكل الانتشار المعطى، كيف يمكننا فعل ذلك؟ بمعنى آخر، ما نوع النموذج الاقتصادي المناسب لوصف مثل هذه العلاقة التي تبدأ منخفضة ثم تتزايد كما هو الحال في التكلفة الحدية؟

هندسيًا، منحني MC الموضح في الشكل (1.7) يمثل parabola يمكن التعبير عنه في المعادلة التالية:

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 \quad (1.10.7)$$

والذي يسمى دالة تربيعية بشكل عام، ويطلق عليه متعددة حدود من الدرجة الثانية في المتغير X – أكبر أس للـ X يمثل درجة متعددة الحدود (إذا كان لدينا X^3 في المعادلة السابقة، فإنها تصبح متعددة حدود من الدرجة الثالثة وهكذا).



شكل (1.7) منحنى التكلفة الحدية الذي يأخذ الشكل U–

الشكل العشوائي لـ (1.10.7) يمكن كتابته كالتالي:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + u_i \quad (2.10.7)$$

ويطلق عليه انحدار متعدد الحدود من الدرجة الثانية. الشكل العام لانحدار متعدد الحدود من الدرجة k يمكن كتابته كالتالي:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \cdots + \beta_k X_i^k + u_i \quad (3.10.7)$$

لاحظ أنه في مثل هذه الانحدارات المتعددة، يوجد لدينا متغير مفسر واحد فقط في الجانب الأيمن من معادلة الانحدار، ولكنه مرفوع لأس مختلف في كل مرة مما

يجعل هذه النماذج نماذج انحدار متعددة. ولاحظ أنه إذا تم اعتبار X ثابتة أو غير عشوائية، فإن مقادير الـ X المرفوعة لأس ما ستعتبر أيضًا ثابتة أو غير عشوائية.

هل مثل هذه النماذج تجعل هناك مشاكل خاصة في التقدير؟ فحيث إن متعددة الحدود من الدرجة الثانية (2.10.7) أو متعددة الحدود من الدرجة k (13.10.7) تعتبر معادلة خطية في المعلمات، فإن الـ β's يمكن تقديرها بطريقة OLS العادية أو ML. ولكن ماذا عن مشكلة الارتباط الخطي؟ أليست المتغيرات X's مرتبطة معًا بشكل كبير، حيث إنها جميعًا عبارة عن نفس المتغير X مرفوعة إلى أس مختلف في كل مرة؟ نعم ولكن تذكر أن المقادير X^2، X^3، X^4 وهكذا جميعها دوال غير خطية في X، وبالتالي وبشكل محدد لا تتعارض مع فرض عدم وجود ارتباط خطي متعدد.

باختصار، فإن نماذج الانحدار المتعددة الحدود يمكن تقديرها باستخدام الأساليب المشروحة في هذا الفصل، وبدون أي مشاكل جديدة في التقدير.

مثال 4.7

تقدير دالة التكلفة:

Estimating the total cost function

كمثال للانحدار المتعدد الحدود، دعنا نستعرض البيانات الخاصة بالإنتاج ودالة التكلفة الكلية لسلعة ما في المدى القصير والمعطاة في الجدول (4.7). ما نوع نموذج الانحدار المناسب لمثل هذه البيانات؟ للإجابة على ذلك، دعنا أولاً نستعرض شكل الانتشار المعطى في الشكل (2.7).

من هذا الشكل واضح أن العلاقة بين التكلفة الكلية والناتج تأخذ الشكل S ولاحظ كيف أن دالة التكلفة الكلية تتزايد تدريجيًا أولاً ثم تتزايد بسرعة كما هو معروف من فكرة العائد المتناقص شكل الـ S الخاص بمنحنى التكلفة الكلية يمكن تمثيله بمتعددة الحدود من الدرجة الثالثة التالية:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \quad (4.10.7)$$

حيث Y = التكلفة الكلية و X = الناتج.

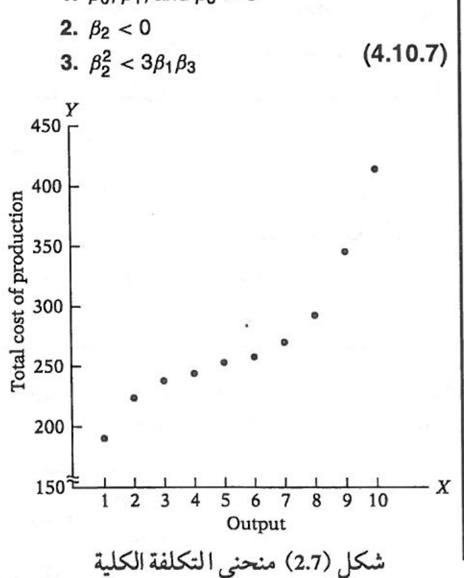
جدول (4.7) التكلفة الكلية (Y) والناتج (X)

Output	Total cost, $
1	193
2	226
3	240
4	244
5	257
6	260
7	274
8	297
9	350
10	420

باستخدام بيانات جدول (4.7)، يمكن تطبيق طريقة OLS للحصول على مقدرات معالم (4.10.7). ولكن قبل القيام بذلك، دعنا نبحث أولاً فيما تفترضه النظرية الاقتصادية عن دالة التكلفة التكعيبية في المدى القصير. نظرية مرونة الأسعار تنص على أنه في المدى القصير،

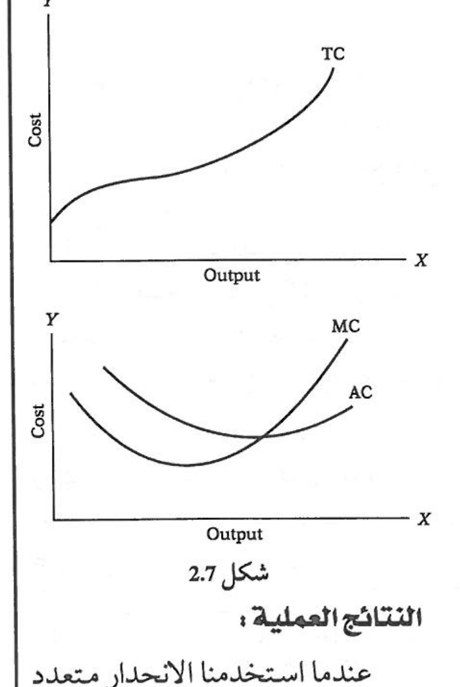
منحنيات التكلفة الحدية (MC) والتكلفة المتوسطة (AC) للإنتاج تكون على الشكل U – فمبدئيًا كلما زاد الناتج فإن الـ MC والـ AC تقل، ولكن بعد الوصول إلى مستوى معين من الناتج، فإن كلهما يبدأ في الزيادة ومرة أخرى ذلك يمثل فكرة العائد المتناقص هذا يمكن رؤيته بوضوح من الشكل (3.7) (انظر أيضًا الشكل 1.7). وبما أن منحنيات MC و AC مشتقة من منحنى التكلفة الكلية، فإن طبيعة الشكل U لهذه المنحنيات يفرض بعض القيود على معلمات منحنى التكلفة الكلية (4.10.7). ففي واقع الأمر، يمكن إثبات أن معالم (4.10.7) يجب أن تستوفي الشروط التالية إذا كانت منحنيات التكلفة الحدية والتكلفة المتوسطة في المدى القصير لها الشكل U وهذه الشروط هي:[19]

$$
\begin{aligned}
&\textbf{1.}\ \beta_0, \beta_1, \text{and } \beta_3 > 0 \\
&\textbf{2.}\ \beta_2 < 0 \\
&\textbf{3.}\ \beta_2^2 < 3\beta_1\beta_3
\end{aligned}
\tag{4.10.7}
$$



شكل (2.7) منحنى التكلفة الكلية

قد تبدو هذه المناقشة النظرية بعيدة عن موضعنا التطبيقي ولكن في واقع الأمر هذه المعرفة النظرية مهمة جداً لاختيار النتائج التطبيقية، حيث إنه إذا لم تكن النتائج العملية متوافقة مع توقعات مسبقة موجودة لدينا فإنه قد نكون وقعنا في خطأ التوصيف (أي اختيار نموذج خاطئ)، فلابد من تعديل الفكرة النظرية أولاً أو البحث عن فكرة جديدة ثم نبدأ التطبيق مرة أخرى. ولكن كما سبق وذكرنا في المقدمة، فإن هذه هي طبيعة البحث التطبيقي.



شكل 2.7

النتائج العملية:

عندما استخدمنا الانحدار متعدد الحدود من الدرجة الثالثة لبيانات جدول (4.7) حصلنا على النتائج التالية:

(19) انظر Alpha C. Chiang, Fundamental Methods of Mathematical Economics, 3d ed., McGraw-Hill, New York, 1984, pp. 250-252.

$$\hat{Y}_i = 141.7667 + 63.4776X_i - 12.9615X_i^2 + 0.9396X_i^3$$
$$(6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591)$$
$$R^2 = 0.9983 \quad (7.10.6)$$

(لاحظ أن: الأرقام الموجودة بين أقواس تمثل الأخطاء القياسية المقدرة). وعلى الرغم من أننا سنختبر المعنوية الإحصائية لهذه النتائج في الفصل التالي، إلا أن القارئ يمكنه إثبات تماشي هذه النتائج مع التوقعات النظرية المنصوص عليها في (6.1.7). ويترك للقارئ كتمرين تفسير وتحليل نتائج هذا الانحدار (6.10.7).

مثال 5.7

معدل نمو GDP، 1960–1985 ومعدل GDP النسبي للفرد في 119 دولة نامية دعنا نستعرض نتائج الانحدار التالي كمثال اقتصادي إضافي لنموذج الانحدار المتعدد الحدود:[20]

$$\widehat{GDPG}_i = 0.013 + 0.062\,RGDP - 0.061\,RGDP^2$$
$$se = (0.004) \quad (0.027) \quad (0.033) \quad (7.10.7)$$
$$R^2 = 0.053 \quad adj\,R^2 = 0.036$$

حيث GDPD = معدل نمو GDP، معطى كنسبة (متوسط في الفترة 1960–1985) وRGDP = GDP النسبي للفرد، 1960 (نسبة لـ GDP الولايات المتحدة بالنسبة للفرد 1960) r2 المعدلة (adj R^2) تعني، مع وضع في الاعتبار عدد المتغيرات المنحدرة)، أن النموذج يفسر حوالي 3.6% من التباين في الـ GDPG. حتى R^2 غير المعدلة 0.053 تبدو قيمة صغيرة. وهذا قد يعتبر ضد جودة النموذج، ولكن كما سنرى في الفصل القادم، قيمة R^2 الصغيرة تظهر كثيراً في البيانات المقطعية التي تشتمل على عدد كبير من المشاهدات، بالإضافة لذلك على الرغم من صفر R^2 فإنها قد تكون معنوية إحصائيًا (بمعنى أنها مختلفة عن الصفر) كما سنرى في الفصل القادم.

كما نرى من هذا الانحدار، GDPG في الدول النامية يتزايد كلما تزايد RGDP ولكن بمعدل متناقص، بمعنى أنه في الدول النامية لا يتم التجاوب بسرعة مع التطور الاقتصادي[21]. هذا المثال يوضح كيف أن نماذج اقتصادية بسيطة يمكن أن تلقي الضوء على ظواهر اقتصادية مهمة.

(20) المصدر: The East Asian Economic Miracle: Economic Growth and Public Policy, A World Bank Policy Research Report, Oxford University Press, U. K, 1993, p. 29.

(21) إذا اشتققنا المعادلة (7.10.7) للحصول على التفاصيل الأولى، نجد أن:

$$\frac{dGDPG}{dRGDP} = 0.062 - 0.122\,RGDP$$

بمعنى أن معدل التغير في GDPG بالنسبة لـ RGDP متناقص. إذا ساوينا ذلك مع الصفر، سنحصل على RDPG≈0.582. أي أنه إذا وصل GDP دولة ما إلى 51% من GDP الولايات المتحدة الأمريكية فإن معدل نمو الـ GDPG سيذهب إلى الصفر.

11.7[*] معاملات الارتباط الجزيئية :
*PARTIAL CORRELATION COEFFIICIENTS

تفسير معاملات الارتباط البسيطة والجزيئية :
Explanation of Simple and Partial Correlation Coefficients

في الفصل (3)، استعرضنا معامل الارتباط r كمقياس لدرجة الارتباط الخطية بين متغيرين اثنين. إذا كان لدينا ثلاثة متغيرات في نموذج الانحدار، فإنه يمكننا حساب ثلاثة معاملات ارتباط: r_{12} (ارتباط بين Y و X_2)، r_{13} (ارتباط بين Y وX_3) وr_{23} (ارتباط بين X_2 و X_3)، لاحظ أننا نستخدم الترميز 1 لـY لسهولة التعريف. معاملات الارتباط هذه تسمى معاملات ارتباط بسيطة أو كلية أو معاملات ارتباط من الدرجة صفر. ويمكن حسابها من خلال تعريف معامل الارتباط المعطى في (13.5.3).

ولكن الآن دعنا نطرح هذا السؤال: هل r_{12} في الحقيقة يقيس درجة الارتباط (الخطية) «الحقيقية» بين Y وX_2 مع الوضع في الاعتبار أن المتغير الثالث X_3 قد يكون مرتبطًا مع أحدهما أو كليهما؟ للإجابة على هذا التساؤل دعنا نستعرض التالي: افترض أن نموذج الانحدار الحقيقي هو (1.1.7) ولكننا حذفنا من النموذج المتغير X_3، وقمنا بعمل انحدار بسيط لـY على X_2 وحصلنا على معامل الميل، مثلاً β_{12}. هل سيكون هذا المعامل مساويًا للمعامل الحقيقي β_2 إذا قمنا باستخدام (7.1.1) للتقدير؟. الإجابة يجب أن تكون واضحة من مناقشتنا للفقرة 7.7. عمومًا، r_{12} في الأغلب لن يمثل درجة الارتباط الحقيقية بين Y وX_2 في وجود X_3. في حقيقة الأمر، سيعطي انطباعًا خاطئًا عن طبيعة العلاقة بين Y و X_2 كما سنرى لاحقًا. وبالتالي ما نحتاج إليه هو معامل ارتباط مستقل عن تأثير X_3 على X_2 أو Y أو كلاهما (إذا كان هناك تأثير). مثل هذا المعامل يمكن الحصول عليه من المعامل المعروف باسم معامل الارتباط الجزيئي. مفهوم هذا المعامل متماثل مع معامل الانحدار الجزيئي، دعنا نعرف الآن التالي:

$r_{12.3}$ = معامل الارتباط الجزيئي بين Y و X_2، مع افتراض ثبات X_3.

$r_{13.2}$ = معامل الارتباط الجزيئي بين Y و X_3، مع افتراض ثبات X_2.

$r_{23.1}$ = معامل الارتباط الجزيئي بين X_2 و X_3، مع افتراض ثبات Y.

(*) اختياري.

معاملات الارتباط الجزيئية هذه يمكن بسهولة الحصول عليها من معاملات الارتباط البسيطة (من الدرجة صفر) كالتالي (للإثبات، انظر التمارين)[22].

$$r_{12.3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}} \quad (1.11.7)$$

$$r_{13.2} = \frac{r_{13} - r_{12}r_{23}}{\sqrt{(1 - r_{12}^2)(1 - r_{23}^2)}} \quad (2.11.7)$$

$$r_{23.1} = \frac{r_{23} - r_{12}r_{13}}{\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}} \quad (3.11.7)$$

معاملات الارتباط الجزيئية المعطاة في المعادلتين (1.11.7) إلى (3.11.7) تسمى معاملات ارتباط من الدرجة الثانية، و $r_{12.345}$ يعتبر معامل ارتباط من الدرجة الثالثة وهكذا. كما لاحظنا من قبل فإن r_{12}، r_{13} ومثيلهما يطلق عليه معاملات ارتباط بسيطة أو من الدرجة صفر. تفسير $r_{12.34}$ هو كالتالي: معامل الارتباط بين Y و X_2 بافتراض ثبات كل من X_3 و X_4.

تفسير معاملات الارتباط الجزيئية والبسيطة:

Interpretation of Simple and Partial Correlation Coefficients

في حالة وجود متغيرين اثنين فقط من نموذج الانحدار، يكون معامل r البسيط من السهل تفسير معناه، فهو يقيس درجة الارتباط (وليس السببية) (الخطية) بين المتغير التابع Y والمتغير المفسر الوحيد X. ولكن بمجرد وجود أكثر من متغيرين، لابد من الحذر عن تفسير معامل الارتباط البسيط. من (1.11.7)، على سبيل المثال، نلاحظ التالي:

1 - حتى اذا كان $r_{12}=0$، فإن $r_{12.3}$ لن يساوى العنصر إلا إذا كان كل من r_{13} و r_{23} كلاهما يساوي الصفر.

2 - إذا كان $r_{12}=0$ و r_{13} و r_{23} لا يساويان الصفر ولهما نفس الإشارة، فإن $r_{12.3}$ ستكون له إشارة سالبة، في حين إذا كانوا مختلفين في الإشارة، فإن $r_{12.3}$ سيكون موجبًا.

دعنا نستعرض المثال التالي لتوضيح هذه النقطة. دع Y = ناتج المحصول، X_2 = الأمطار و X_3 = درجة الحرارة. بافتراض أن $r_{12}=0$ فإن ذلك يعني عدم

(22) معظم برامج الحاسب الآلي الخاصة بتحليل الانحدار المتعدد تحسب معاملات الارتباط البسيطة وبالتالي يكون من السهل حساب معاملات الارتباط الجزيئية.

وجود ارتباط بين ناتج المحصول والأمطار. افترض أيضًا أن r_{13} موجب و r_{23} سالب بالتالي من (1.11.7) سيكون $r_{12.3}$ موجبًا، بمعنى أنه بافتراض ثبات درجة الحرارة، فإن هناك ارتباطًا طرديًا (موجبًا) بين الناتج والأمطار. وهنا يتضح تعارض النتائج، عمومًا، ذلك غير مستبعد. حيث إن درجة الحرارة X_3 تؤثر على كل من Y والأمطار X_2، وبالتالي لإيجاد العلاقة الخالصة بين ناتج المحصول والامطار، لابد من إزالة أثر متغير الحرارة. وهذا المثال يوضح لنا كيف يمكن أن تقع في خطأ عدم الفهم الصحيح للعلاقة بسبب استخدامنا لمعامل الارتباط البسيط.

3 - المقادير $r_{12.3}$ و r_{12} (ومثيلهما) قد لا يكون لهما نفس الإشارة.

4 - في حالة وجود متغيرين اثنين، رأينا أن r_2 يكون بين 0 و1. نفس هذه الخاصية متحققة لمربع معاملات الارتباط الجزيئية. وبناء على ذلك، يمكن للقارئ إثبات أنه يمكن الحصول على المعادلة التالية من المعادلة (1.11.7):

$$0 \leq r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} \leq 1 \tag{4.11.7}$$

وهي تمثل العلاقة المتداخلة بين معاملات الارتباط الثلاثة من الدرجة صفر.

معاملات مماثلة يمكن الحصول عليها من المعادلات (3.9.7) و (4.9.7).

5 - افترض أن $r_{13} = r_{23} = 0$. هل يعني ذلك أن r_{12} أيضًا يساوي الصفر؟ الإجابة واضحة من (4.11.7). حقيقة أن Y و X_3 من جهة و Y_2 و X_3 من جهة أخرى غير مرتبطين لا تعني على الإطلاق أن Y، X_2 غير مرتبطين.

وعمومًا لاحظ أن $r_{12.3}^2$ يطلق عليه أحيانًا معامل التحديد الجزيئي، ويمكن أن يفسر على أنه نسبة التباين في Y غير المفسرة من خلال المتغير X_3، ولكن المفسرة من خلال وجود X_3 في النموذج (انظر تمرين 5.7). وبالتالي تعريفيًا هي مماثلة لـ R^2.

وقبل التطرق إلى موضوع آخر، لاحظ العلاقات التالية بين R^2، معاملات الارتباط البسيطة ومعاملات الارتباط الجزيئية.

$$R^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2} \tag{5.11.7}$$

$$R^2 = r_{12}^2 + \left(1 - r_{12}^2\right) r_{13.2}^2 \tag{6.11.7}$$

$$R^2 = r_{13}^2 + \left(1 - r_{13}^2\right) r_{12.3}^2 \tag{7.11.7}$$

وفي نهاية هذه الفقرة، دعنا نستعرض الاستنتاج التالي: سبق وأن ذكرنا أن R^2 لن تقل إذا أضفنا متغيراً مفسراً جديداً للنموذج، وذلك يتضح من (6.11.7). فهذه المعادلة تعني أن نسبة تباين الـ Y المفسرة من خلال X_2 و X_3 معًا هى مجموع جزءين: جزء خاص بتغير مرتبط بـ X_2 بمفردة ($= r_{12}^2$) وجزء غير مرتبط بـ X_2 ($=1-r_{12}^2$) مضروب في النسبة المفسرة من خلال X_3 بعد افتراض ثبات X_2. الآن $R^2 > r_{12}^2$ طالمًا أن $r_{13.2}^2 > 0$. فعلى أسوأ الأحوال فإن $r_{13.2}^2$ سيساوي الصفر إذا كان $R^2 = r_{12}^2$.

12.7 الخلاصة والنتائج : SUMMARY AND CONCLUSIONS

1 – قدمنا في هذا الفصل، نموذج الانحدار الخطي المتعدد الأكثر بساطة، وهو النموذج ثلاثي المتغيرات. وأوضحنا أن المصطلح (الخطي) يعني أنه نموذج خطي في المعلمات وليس بالضرورة أن يكون خطيًا في المتغيرات.

2 – على الرغم من أن النموذج ثلاثي المتغيرات، هو في أغلب الأحيان يكون امتداداً للنموذج ثنائي المتغيرات، إلا أن هناك بعض المفاهيم الجديدة الخاصة به مثل معاملات الانحدار الجزئية، معاملات الارتباط الجزئية، معاملات الارتباط المتعددة، R^2 المعدلة وغير المعدلة (مع درجات الحرية)، الارتباط الخطي المتعدد وتحيز التوصيف.

3 – قدمنا في هذا الفصل أيضًا أشكال الدوال المختلفة لنموذج الانحدار المتعدد، مثل دالة إنتاج Cobb Dauglas ونموذج الانحدار المتعدد الحدود.

4 – وعلى الرغم من أن R^2 و R^2 المعدلة يقيسان عمومًا كيف أن النموذج مناسب بالنسبة للبيانات، إلا أنه يجب عدم تضخيم الدور الذي يقومان به، المهم والضروري هو التوقعات النظرية الكامنة وراء النموذج بمعنى الإثارة المتوقعة لمعاملات المتغيرات الموجودة في النموذج ومعنوياتهم الإحصائية، كما سيتم شرحها تفصيلاً في الفصل القادم.

5 – النتائج المقدمة في هذا الفصل، من الممكن تعميمها لأي نموذج انحدار خطي متعدد يشمل على أي عدد من المتغيرات المنحدرة. ولكن العمليات الجبرية ستكون معقدة نوعًا ما. ولكن هذا التعقيد سيختفى تمامًا إذا تناولنا هذه العمليات في صورة مصفوفات.. وللقارئ الذي يرغب في معرفة المزيد عن

ذلك، فإن نموذج الانحدار الذي يشمل على k متغير مشروع باستخدام جبر المصفوفات في الملحق 2، وهذا الجزء اختياري لمن يرغب في قراءته. ولكن لعامة القراء، فإنه يمكن تكملة الموضوع بدون معرفة الكثير عن جبر المصفوفات.

تمارين : EXERCISES

أسئلة : Questions

1.7 اعتبر البيانات التالية في جدول (5.7).

جدول (5.7)

Y	X_2	X_3
1	1	2
3	2	1
8	3	−3

بناء على هذه البيانات، قدر الانحدارات التالية:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i} \quad (1)$$

$$Y_i = \lambda_1 + \lambda_3 X_{3i} + u_{2i} \quad (2)$$

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \quad (3)$$

ملحوظة: قدر معاملات الانحدار فقط وليس الأخطاء القياسية.

(a) هل $\beta_2 = \alpha_2$؟ علل إجابتك.

(b) هل $\lambda_3 = \beta_3$؟ علل إجابتك.

ما النتيجة المهمة التي توصلت إليها من خلال هذا التمرين؟

2.7 من البيانات التالية، قدر معاملات الانحدار الجزيئية، أخطاؤهم القياسية وقيمة R^2 المعدلة وغير المعدلة.

$$\bar{Y} = 367.693 \qquad \bar{X}_2 = 402.760 \qquad \bar{X}_3 = 8.0$$

$$\sum (Y_i - \bar{Y})^2 = 66042.269 \qquad \sum (X_{2i} - \bar{X}_2)^2 = 84855.096$$

$$\sum (X_{3i} - \bar{X}_3)^2 = 280.000 \qquad \sum (Y_i - \bar{Y})(X_{2i} - \bar{X}_2) = 74778.346$$

$$\sum (Y_i - \bar{Y})(X_{3i} - \bar{X}_3) = 4250.900 \qquad \sum (X_{2i} - \bar{X}_2)(X_{3i} - \bar{X}_3) = 4796.000$$

$$n = 15$$

3.7 اثبت أن المعادلة (7.4.7) يمكن كتابتها كالتالي:

$$\hat{\beta}_2 = \frac{\sum y_i(x_{2i} - b_{23}x_{3i})}{\sum(x_{2i} - b_{23}x_{3i})^2}$$

$$= \frac{\text{net (of } x_3\text{) covariation between } y \text{ and } x_2}{\text{net (of } x_3\text{) variation in } x_2}$$

حيث b_{23} هو معامل الميل في انحدار X_2 على X_3

(ملحوظة: اجعل $b_{23} = \frac{\sum x_{2i}x_{3i}}{\sum x_{3i}^2}$)

4.7 في نموذج انحدار متعدد وجدت أن مقدار الخطأ u_i له التوزيع التالي، $u_i \sim N(0, 4)$. كيف يمكن لك تصميم تجربة Monte Carlo لإثبات أن التباين الحقيقي يساوي 4.

5.7 اثبت أن $r^2_{12.3} = (R^2 - r^2_{13})/(1 - r^2_{13})$ وفسر المعادلة.

6.7 إذا كان $\alpha_1X_1 + \alpha_2X_2 + \alpha_3X_3 = 0$ لكل قيم X_1، X_2، X_3. اوجد قيم معاملات الارتباط الجزيئية الثلاثة.

7.7 هل يمكن الحصول على التالي من أي بيانات؟

(a) $r_{23} = 0.9$ ، $r_{13} = -0.2$ ، $r_{12} = 0.8$

(b) $r_{12} = 0.6$ ، $r_{23} = -0.9$ ، $r_{31} = -0.5$

(c) $r_{21} = 0.01$ ، $r_{13} = 0.66$ ، $r_{23} = -0.7$

8.7 اعتبر النموذج التالي:

$Y_i = \beta_1 + \beta_2$ التعليم $+ \beta_3$ عدد سنوات الخبرة $+u_i$

بافتراض حذف متغير عدد سنوات الخبرة. ما نوعية المشاكل والتحيز الذي قد تواجه؟ فسر لغويًا.

9.7 أثبت أن β_2 و β_3 في (2.9.7)، يعطيان فعليًا مرونة العمالة ورأس المال. (هذا السؤال يمكن الإجابة عليه بدون أي حسابات، فقط باستخدام تعريف معامل المرونة، وتذكر أن التغير في لوغاريتم المتغير هو تغير نسبي بافتراض أن التغيرات صغيرة نوعًا ما).

10.7 بافتراض نموذج الانحدار الخطي ثلاثي المتغيرات الذي ناقشناه في هذا الفصل:

(a) افترض أننا ضربنا كل قيم X_2 في 2. ما أثر ذلك، إذا كان هناك أي أثر، على تقديرات المعاملات وأخطائهم القياسية؟

(b) بدلاً من a، افترض الآن أننا ضربنا كل قيم Y في 2. ما أثر ذلك، إذا كان هناك أي أثر، على تقديرات المعاملات وأخطائهم القياسية؟

11.7 عمومًا $R^2 \neq r_{12}^2 + r_{13}^2$ ولكن ذلك يتحقق فقط إذا كان $r_{23} = 0$ علق على ذلك ووضح مدى معنوية ذلك [انظر المعادلة (5.11.7)].

12.7 اعتبر النماذج التالية: (*)

نموذج A: $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_{1t}$

نموذج B: $(Y_t - X_{2t}) = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_{2t}$

(a) هل مقدرات OLS لـ α_1 و β_1 ستكون متساوية؟ لماذا؟

(b) هل مقدرات OLS لكل من α_3 و β_3 ستكون متساوية؟ لماذا؟

(c) ما هي العلاقة بين α_2 و β_2؟

(d) هل يمكنك مقارنة قيم R^2 للنموذجين؟ علل إجابتك.

13.7 افترض أنك تريد تقدير دالة الاستهلاك التالية(**):

$$Y_i = \alpha_1 + \alpha_2 X_2 + u_{1i}$$

ودالة الادخار التالية:

$$Z_i = \beta_1 + \beta_2 X_i + u_{2i}$$

حيث Y = الاستهلاك، Z = الادخار، X = الدخل و $X = Y + Z$ بمعنى أن الدخل يساوي الاستهلاك مضافًا إليه الادخار.

(a) ما هي العلاقة، إذا كانت هناك علاقة، بين α_2 و β_2؟ وضح إجابتك بالخطوات الحسابية المطلوبة.

(b) هل مجموع مربعات البواقي RSS سيكون متساويًا للنموذجين؟ علل إجابتك.

(*) مأخوذ من: Wojciech W. Charemza and Derek F. Deadman, Econometric Practice: General to Specific Modelling, Cointegration and Vector Autogression, Edward Elgar, Brookfield, Vermont, 1992, p. 18.

(**) مأخوذ من: Peter Kennedy, A Guide to Econometrics, 3d ed., The MIT Press, Cambridge, Massachusetts, 1992, p. 308, Question #9.

(c) هل يمكنك مقارنة قيم R^2 الخاصة بالنموذجين؟ علل إجابتك.

14.7 افترض أن دالة Cobb - Douglas المعطاة في (1.9.7) تم كتابتها كالتالي:

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} u_i$$

إذا أخذنا لوغاريتم الطرفين لهذا النموذج، سيكون لديك $\ln u_i$ كمقدار الخطأ على الجانب الأيمن من المعادلة.

(a) ما هو الفرض الاحتمالي الخاص بـ $\ln u_i$ والضروري لتطبيق نموذج الانحدار الخطي المعتاد التقليدي (CNLRM)؟ كيف يمكنك اختيار ذلك بالنسبة للبيانات المعطاة في جدول (3.7)؟

(b) هل نفس الفروض يجب توافرها في u_i؟ علل إجابتك.

15.7 الانحدار المار بنقطة الأصل. اعتبر نموذج الانحدار التالي المار بنقطة الأصل:

$$Y_i = \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i$$

(a) كيف يمكنك تقدير المعلمات المجهولة؟

(b) هل $\Sigma \hat{u}_i$ سيكون مساويًا للصفر في هذا النموذج؟ علل إجابتك.

(c) هل $\Sigma \hat{u}_i X_{2i} = \Sigma \hat{u}_i X_{3i} = 0$ في هذا النموذج؟

(d) هل يمكنك تعميم نتائج لتشمل حالة نموذج له k متغير؟

(**ملحوظة**: اتبع المناقشة الخاصة بحالة متغيرين اثنين فقط المعطاة في الفصل 6).

مسـائل: Problems

16.7 الطلب على الزهور(*). جدول (6.7) يعطي بيانات ربع سنوية عن المتغيرات التالية:

Y= كمية الزهور المباعة، بالدزينة (الدستة)

X_2 = متوسط سعر الجملة للزهور، الدولار/ دزينة

X_3 = متوسط سعر الجملة، الدولار/ دزينة

X_4 = متوسط دخل الأسرة الأسبوعي، الدولار/ أسبوع

X_5 = متغير الاتجاه ويأخذ القيم 1، 2، وهكذا للفترة

(*) خالص الشكر لـ Joe Walsh الذي جمع البيانات الخاصة بهذا المثال من بائعي الجملة في منطقة العاصمة ديترويت وتتبعه لكل خطوات هذه الدراسة.

III–1971 إلى II–1975 في منطقة العاصمة ديترويت.

والآن مطلوب منك أولاً اعتبار دوال الطلب التالية:

$$Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{4t} + \alpha_5 X_{5t} + u_t$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 X_{5t} + u_t$$

(a) قدر معالم النموذج الخطي – فسر نتائجه.

(b) قدر معالم النموذج اللوغاريتمي الخطي وفسر نتائجه.

(c) β_2، β_3 و β_4 تمثل بالترتيب سعر الملكية، السعر المخالف، ومرونة الدخل بالنسبة للطلب، ما الذي تتوقعه بالنسبة لإشارة كل منها؟ هل النتائج متطابقة مع هذه التوقعات؟

جدول (6.7)

Year and quarter	Y	X_2	X_3	X_4	X_5
1971–III	11,484	2.26	3.49	158.11	1
–IV	9,348	2.54	2.85	173.36	2
1972–I	8,429	3.07	4.06	165.26	3
–II	10,079	2.91	3.64	172.92	4
–III	9,240	2.73	3.21	178.46	5
–IV	8,862	2.77	3.66	198.62	6
1973–I	6,216	3.59	3.76	186.28	7
–II	8,253	3.23	3.49	188.98	8
–III	8,038	2.60	3.13	180.49	9
–IV	7,476	2.89	3.20	183.33	10
1974–I	5,911	3.77	3.65	181.87	11
–II	7,950	3.64	3.60	185.00	12
–III	6,134	2.82	2.94	184.00	13
–IV	5,868	2.96	3.12	188.20	14
1975–I	3,160	4.24	3.58	175.67	15
–II	5,872	3.69	3.53	188.00	16

(d) كيف يمكنك حساب كل من السعر، السعر البديل، ومرونة الدخل لهذا النموذج الخطي؟

(e) بناء على تحليك للبيانات، أي النماذج تختار ولماذا؟

17.7 أنشطة wildcat .wildcats تقوم بأعمال تنقيب للبحث عن آبار بترول جديدة، أما في منطقة جديدة أو في منطقة سبق وأن وجد فيها بترول من قبل، ويتم

التنقيب فيها لزيادة المتوافر منها من مخزون البترول. جدول (7.7) يعطي البيانات الخاصة بهذه المتغيرات(*):

Y = عدد مرات التنقيب

X_2 = السعر في الفترة السابقة (بتثبيت الدولار، 1972 = 100)

X_3 = الإنتاج المحلي

X_4 = GNP (1972 = 100)

X_5 = متغير الاتجاه، 1948 = 1، 1949 = 2، . . . ، 1978 = 31

ادرس مدى توفيق النموذج التالي لتمثيل البيانات:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 \ln X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

(a) هل يمكن تقديم توقع سابق منطقي لهذا النموذج؟

(b) بافتراض قبول النموذج، قدر معالم النموذج وأخطاءه القياسية، واحصل على R^2 و $\bar{R}^2$.

(c) علق على نتائجك وفقًا لتوقعاتك المسبقة.

(d) ما الأشياء الأخرى الممكن اعتيادها لتفسير أنشطة wildcat؟ لماذا؟

جدول (7.7)

Thousands of wildcats, (Y)	Per barrel price, constant $, ($X_2$)	Domestic output (millions of barrels per day), (X_3)	GNP, constant $ billions, (X_4)	Time, (X_5)
8.01	4.89	5.52	487.67	1948 = 1
9.06	4.83	5.05	490.59	1949 = 2
10.31	4.68	5.41	533.55	1950 = 3
11.76	4.42	6.16	576.57	1951 = 4
12.43	4.36	6.26	598.62	1952 = 5
13.31	4.55	6.34	621.77	1953 = 6
13.10	4.66	6.81	613.67	1954 = 7
14.94	4.54	7.15	654.80	1955 = 8
16.17	4.44	7.17	668.84	1956 = 9
14.71	4.75	6.71	681.02	1957 = 10
13.20	4.56	7.05	679.53	1958 = 11
13.19	4.29	7.04	720.53	1959 = 12
11.70	4.19	7.18	736.86	1960 = 13
10.99	4.17	7.33	755.34	1961 = 14

(*) خالص الشكر لـ Raymond Sauino لجمع البيانات والتعامل معها.

1963 = 16	830.70	7.61	4.04	10.66
1964 = 17	874.29	7.80	3.96	10.75
1965 = 18	925.86	8.30	3.85	9.47
1966 = 19	980.98	8.81	3.75	10.31
1967 = 20	1,007.72	8.66	3.69	8.88
1968 = 21	1,051.83	8.78	3.56	8.88
1969 = 22	1,078.76	9.18	3.56	9.70
1970 = 23	1,075.31	9.03	3.48	7.69
1971 = 24	1,107.48	9.00	3.53	6.92
1972 = 25	1,171.10	8.78	3.39	7.54
1973 = 26	1,234.97	8.38	3.68	7.47
1974 = 27	1,217.81	8.01	5.92	8.63
1975 = 28	1,202.36	7.78	6.03	9.21
1976 = 29	1,271.01	7.88	6.12	9.23
1977 = 30	1,332.67	7.88	6.05	9.96
1978 = 31	1,385.10	8.67	5.89	10.78

المصدر : Energy Information Administration, 1978 Report to Congress.

18.7 الميزانية المنفقة على الدفاع في الولايات المتحدة الأمريكية، 1962 – 1981 لتفسير ميزانية U. S. للدفاع، دعنا نعتبر النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + u_t$$

حيث:

Y_t = المنفق من المال على ميزانية الدفاع الأمريكية في السنة t (بلايين الدولارات)

X_{2t} = GNP في السنة t (بلايين الدولارات)

X_{3t} = مبيعات / مساعدات الجيش الأمريكي في السنة t (بلايين الدولارات)

X_{4t} = مبيعات صناعة أسلحة الفضاء، (بلايين الدولارات)

X_{5t} = الجيوش المشتركة في صراع ما، والمشتملة على أكثر من 100,000 جندي، هذا المتغير يأخذ القيمة 1 عندما يكون هناك 100,000 جندي أو أكثر مشتركين في صراع ما ويساوي الصفر عندما يكون عدد الجنود أقل من 100,000.

لاختيار هذا النموذج، لديك البيانات المعطاة في جدول (8.7).

(a) قدر معالم هذا النموذج والأخطاء القياسية لها واحصل على R^2 و R^2 المعدلة و$\bar{R}^2$.

(b) علق على النتائج، مع الوضع في الاعتبار أي توقعات مسبقة عن العلاقة بين Y وكل متغير من المتغيرات المفسرة X.

(c) ما هي المتغيرات الأخرى التي قد ترغب في إضافتها للنموذج، ولماذا؟

19.7 الطلب على الدواجن في الولايات المتحدة الأمريكية، 1960-1982.

لدراسة استهلاك الفرد من الدواجن في الولايات المتحدة، لديك البيانات المعطاة في جدول (9.7) حيث:

Y = الاستهلاك الفردي من الدواجن، 1b.

X_2 = الدخل الجاري الحقيقي الفردي، $.

X_3 = سعر المفرق الحقيقي للدواجن لكل 1b، ¢.

X_4 = سعر المفرق الحقيقي للخنازير لكل 1b، ¢.

X_5 = سعر المفرق الحقيقي للحوم البقر لكل 1b، ¢.

X_6 = سعر المفرق البديل للدواجن لكل 1b، ¢ وهو عبارة عن متوسط مرجح من أسعار المفرق الحقيقية لكل 1b من الخنازير واللحوم البقر، الأوزان عبارة عن الاستهلاك النسبي للحوم البقر أو للخنازير بالنسبة لإجمالي استهلاك للحوم البقر والخنازير.

جدول (8.7)

Year	Defense budget outlays, Y	GNP, X_2	U.S. military sales/ assistance, X_3	Aerospace industry sales, X_4	Conflicts 100,000+, X_5
1962	51.1	560.3	0.6	16.0	0
1963	52.3	590.5	0.9	16.4	0
1964	53.6	632.4	1.1	16.7	0
1965	49.6	684.9	1.4	17.0	1
1966	56.8	749.9	1.6	20.2	1
1967	70.1	793.9	1.0	23.4	1
1968	80.5	865.0	0.8	25.6	1
1969	81.2	931.4	1.5	24.6	1
1970	80.3	992.7	1.0	24.8	1
1971	77.7	1,077.6	1.5	21.7	1
1972	78.3	1,185.9	2.95	21.5	1
1973	74.5	1,326.4	4.8	24.3	0
1974	77.8	1,434.2	10.3	26.8	0
1975	85.6	1,549.2	16.0	29.5	0
1976	89.4	1,718.0	14.7	30.4	0
1977	97.5	1,918.3	8.3	33.3	0
1978	105.2	2,163.9	11.0	38.0	0
1979	117.7	2,417.8	13.0	46.2	0
1980	135.9	2,633.1	15.3	57.6	0
1981	162.1	2,937.7	18.0	68.9	0

المصدر: تم تجميع البيانات عن طريق Albert lucchino من منشورات حكومية متعددة.

جدول (9.7)

Year	Y	X_2	X_3	X_4	X_5	X_6
1960	27.8	397.5	42.2	50.7	78.3	65.8
1961	29.9	413.3	38.1	52.0	79.2	66.9
1962	29.8	439.2	40.3	54.0	79.2	67.8
1963	30.8	459.7	39.5	55.3	79.2	69.6
1964	31.2	492.9	37.3	54.7	77.4	68.7
1965	33.3	528.6	38.1	63.7	80.2	73.6
1966	35.6	560.3	39.3	69.8	80.4	76.3
1967	36.4	624.6	37.8	65.9	83.9	77.2
1968	36.7	666.4	38.4	64.5	85.5	78.1
1969	38.4	717.8	40.1	70.0	93.7	84.7
1970	40.4	768.2	38.6	73.2	106.1	93.3
1971	40.3	843.3	39.8	67.8	104.8	89.7
1972	41.8	911.6	39.7	79.1	114.0	100.7
1973	40.4	931.1	52.1	95.4	124.1	113.5
1974	40.7	1,021.5	48.9	94.2	127.6	115.3
1975	40.1	1,165.9	58.3	123.5	142.9	136.7
1976	42.7	1,349.6	57.9	129.9	143.6	139.2
1977	44.1	1,449.4	56.5	117.6	139.2	132.0
1978	46.7	1,575.5	63.7	130.9	165.5	132.1
1979	50.6	1,759.1	61.6	129.8	203.3	154.4
1980	50.1	1,994.2	58.9	128.0	219.6	174.9
1981	51.7	2,258.1	66.4	141.0	221.6	180.8
1982	52.9	2,478.7	70.4	168.2	232.6	189.4

المصدر: بيانات الـ Y مأخوذة من Gitibase وبالنسبة لـ X_2 حتى X_6 من قطاع الزراعة الأمريكي.

خالص الشكر لـ Robert J. Fisher على تجميع البيانات وتحليلها إحصائيًا.

لاحظ أن: الأسعار الحقيقية تم الحصول عليها باسم الأسعار الاسمية على مؤشر سعر الاستهلاك للطعام.

والآن دعنا نعتبر دوال الطلب التالية:

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + u_t \quad (1)$$

$$\ln Y_t = \gamma_1 + \gamma_2 \ln X_{2t} + \gamma_3 \ln X_{3t} + \gamma_4 \ln X_{4t} + u_t \quad (2)$$

$$\ln Y_t = \lambda_1 + \lambda_2 \ln X_{2t} + \lambda_3 \ln X_{3t} + \lambda_4 \ln X_{5t} + u_t \quad (3)$$

$$\ln Y_t = \theta_1 + \theta_2 \ln X_{2t} + \theta_3 \ln X_{3t} + \theta_4 \ln X_{4t} + \theta_5 \ln X_{5t} + u_t \quad (4)$$

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{6t} + u_t \quad (5)$$

من نظرية الاقتصاد الجزيئي، معروف أن الطلب على سلعة ما عمومًا يعتمد على الدخل الحقيقي للمستهلك، السعر الحقيقي للسلعة، والأسعار الحقيقية للسلع المنافسة والمكملة.. باعتبار كل ذلك، اجب عن الأسئلة الآتية:

(a) أي من دوال الطلب السابقة يمكنك اختياره ولماذا؟

(b) كيف يمكنك تفسير معاملات $\ln X_{2t}$ و $\ln X_{3t}$ في هذه النماذج؟

(c) ما هو الفرق بين التوصيف (2) و (4)؟

(d) ما المشاكل التي ستواجهها إذا أخذت النموذج (4)؟

(**لاحظ أن**: أسعار الخنازير ولحوم البقر موجودة مع أسعار الدواجن).

(e) بما أن التوصيف (5) عن دالة (4)؟ لماذا؟

(f) هل الخنازير ولحوم البقر تعتبر سلعًا بديلة أو تنافسية للدواجن؟ كيف عرفت ذلك؟

(g) افترض أن الدالة (5) هي دالة الطلب «الصحيحة» قدر معالم هذا النموذج، احصل على الأخطاء القياسية لها و R^2 و$\bar{R}^2$ و R^2 المعدلة فسر نتائجك.

(h) الآن افترض أنك قمت بإجراء النموذج «غير الصحيح» رقم (2). حدد عواقب هذا التوصيف الخاطئ مع اعتبار قيم γ_2 و γ_3 وعلاقتها بـ β_2 و β_3 بالترتيب (**ملحوظة**: ارجع إلى الفقرة 7.7).

20.7 في دراسة عن تحول سوق العمالة (James F. Ragon, Jr.) حصلنا على النتائج التالية الخاصة بالاقتصاد الأمريكي في الفترة من I–1950 إلى(*) IV–1979 (الأرقام بين الأقواس تمثل تقدير إحصاء t).

$$\begin{aligned} \widehat{\ln Y_t} = \ & 4.47 \ - \ 0.34 \ln X_{2t} + 1.22 \ln X_{3t} + 1.22 \ln X_{4t} \\ & (4.28) \quad (-5.31) \qquad (3.64) \qquad (3.10) \\ & + 0.80 \ln X_{5t} - 0.0055 X_{6t} \qquad \bar{R}^2 = 0.5370 \\ & \quad (1.10) \qquad (-3.09) \end{aligned}$$

ملحوظة: سنناقش الإحصاء t في الفصل التالي.

حيث:

Y : معدل ترك العمل في المصانع، يعرف على أنه عدد الأفراد الذين يتركون العمل برغبتهم المطلقة لكل 100 عامل.

X_2 = متغير اصطناعي لمعدل بطالة الذكور البالغين.

X_3 = نسبة العمالة أقل من 25.

$X_4 = \frac{N_{t-1}}{N_{t-4}}$ = نسبة العمالة في الربع $(t-1)$ إلى الربع الأخير من السنة $(t-1)$

(*) المصدر: انظر Ragan's article, "Turnover in the Labor Market: A Study of Quit and Layoff Rates," Economic Review, Federal Reserve Bank of Kansas City, May 1981, pp. 13–22.

X_5 = نسبة عمالة المرأة.

X_6 = متغير الزمن (الاتجاه) (1950–I=1)

(a) فسر النتائج المعطاة.

(b) هل العلاقة العكسية الموجودة بين لوغاريتم Y و X_2 لها توقع مسبق لديك؟

(c) لماذا يأخذ معامل $\ln X_3$ إشارة موجبة؟

(d) بما أن معامل الاتجاه سالب. فهناك معدل متناقص مع مرور الزمن. ما نسبة ذلك في معدل ترك العمل؟ ولماذا يوجد مثل هذا التناقص؟

(e) هل $\bar{R}^2$ منخفضة أكثر من اللازم؟

(f) هل يمكنك تقدير الأخطاء القياسية لمعاملات الانحدار من البيانات المعطاة؟ علل إجابتك.

21.7 اعتبر دالة الطلب على المال التالية في الولايات المتحدة الأمريكية في الفترة من 1980 إلى 1998:

$$M_t = \beta_1 Y_t^{\beta_2} r_t^{\beta_3} e^{u_t}$$

حيث:

M = الطلب الحقيقي على المال، باستخدام تعريف M2 للمال.

Y = GDP الحقيقي.

r = معدل الفائدة.

لتقدير دالة الطلب على المال السابقة، لديك البيانات اللازمة في جدول (10.7).

ملاحظة: لتحويل الكميات الاسمية إلى كميات حقيقية، اقسم M و GDP على CPI. ولا يوجد احتياج لقسمة معدل الفائدة على CPT.

ولاحظ أيضًا أن لديك معدلين للفائدة. معدل قصير المدى مقاس لمعدل كل ثلاثة شهور، ومعدل طويل المدى مقاس لسندات لمدة 30 عامًا، ومسبقًا هناك العديد من الدراسات التطبيقية التي استخدمت كلاً من هذين النوعين من معدلات الفائدة.

(a) باستخدام البيانات قدر دالة الطلب السابقة. ما هي مرونة الدخل، ومرونة معدل الفائدة للطلب على المال؟

(b) بدلاً من تقدير دالة الطلب السابقة، افترض أنك تريد استخدام الدالة $(M/Y)_t = \alpha_1 r_t^{\alpha 2} e^{ut}$. كيف يمكنك تفسير النتائج؟ وضح الخطوات الحسابية الرئيسية لذلك.

(c) كيف يمكنك اختيار النموذج الأفضل؟ (ملاحظة: هناك اختبار إحصائي لذلك سيتم تناوله تفصيلاً في الفصل 8).

22.7 جدول (11.7) يمثل بيانات قطاع الصناعة في الاقتصاد اليوناني خلال الفترة 1961–1987.

جدول (10.7) الطلب على المال في الولايات المتحدة الأمريكية، 1980–1998

Observation	GDP	M2	CPI	LTRATE	TBRATE
1980	2795.6	1600.4	82.4	11.27	11.506
1981	3131.3	1756.1	90.9	13.45	14.029
1982	3259.2	1911.2	96.5	12.76	10.686
1983	3534.9	2127.8	99.6	11.18	8.630
1984	3932.7	2311.7	103.9	12.41	9.580
1985	4213.0	2497.4	107.6	10.79	7.480
1986	4452.9	2734.0	109.6	7.78	5.980
1987	4742.5	2832.8	113.6	8.59	5.820
1988	5108.3	2995.8	118.3	8.96	6.690
1989	5489.1	3159.9	124.0	8.45	8.120
1990	5803.2	3279.1	130.7	8.61	7.510
1991	5986.2	3379.8	136.2	8.14	5.420
1992	6318.9	3434.1	140.3	7.67	3.450
1993	6642.3	3487.5	144.5	6.59	3.020
1994	7054.3	3502.2	148.2	7.37	4.290
1995	7400.5	3649.3	152.4	6.88	5.510
1996	7813.2	3824.2	156.9	6.71	5.020
1997	8300.8	4046.7	160.5	6.61	5.070
1998	8759.9	4401.4	163.0	5.58	4.810

ملاحظات: GDP: إجمالي الناتج المحلي (بلايين $).

M2: عرض المال M2

CPI: مؤشر سعر الاستهلاك (100 = 1982 – 1984)

LTRATE: معدل الفائدة في المدى الطويل (سندات 30 عامًا)

TBRATE: معدل الفائدة في المدى القصير (3 شهور، % لكل سنة)

المصدر: Elononic Report of th president, 2000, tables B-1, B-5, B-67, B-71.

جدول (11.7) قطاع الصناعة اليوناني

Observation	Output*	Capital	Labor†	Capital-to-labor ratio
1961	35.858	59.600	637.0	0.0936
1962	37.504	64.200	643.2	0.0998
1963	40.378	68.800	651.0	0.1057
1964	46.147	75.500	685.7	0.1101
1965	51.047	84.400	710.7	0.1188
1966	53.871	91.800	724.3	0.1267
1967	56.834	99.900	735.2	0.1359
1968	65.439	109.100	760.3	0.1435
1969	74.939	120.700	777.6	0.1552
1970	80.976	132.000	780.8	0.1691
1971	90.802	146.600	825.8	0.1775
1972	101.955	162.700	864.1	0.1883
1973	114.367	180.600	894.2	0.2020
1974	101.823	197.100	891.2	0.2212
1975	107.572	209.600	887.5	0.2362
1976	117.600	221.900	892.3	0.2487
1977	123.224	232.500	930.1	0.2500
1978	130.971	243.500	969.9	0.2511
1979	138.842	257.700	1006.9	0.2559
1980	135.486	274.400	1020.9	0.2688
1981	133.441	289.500	1017.1	0.2846
1982	130.388	301.900	1016.1	0.2971
1983	130.615	314.900	1008.1	0.3124
1984	132.244	327.700	985.1	0.3327
1985	137.318	339.400	977.1	0.3474
1986	137.468	349.492	1007.2	0.3470
1987	135.750	358.231	1000.0	0.3582

* بلايين الدراخم عند أسعار 1970 الثابتة.

† آلاف العمال في العام الواحد

المصدر: خالص الشكر لـ George K. Zestos من جامعة Christopher Newport، Virgirer، لمساعدتي بهذه البيانات.

(a) ادرس ما إذا كانت دالة إنتاج Cobb-douglas مناسبة للبيانات المعطاة في الجدول وفسر النتائج. ما هو الاستنتاج العام الذي توصلت إليه؟

(b) والآن دعنا نعتبر النموذج التالي:

$$\text{output/labor} = A(K/L)^{\beta}e^{u}$$

حيث إن المتغير المنحدر عليه هو إنتاجية العمالة، والمتغير المنحدر يمثل نسبة رأس المال للعمالة. ما هي المعنوية الاقتصادية لمثل هذه العلاقة، إذا كانت هناك أي معنوية لها؟ قدر معالم هذا النموذج وفسر نتائجك.

23.7 بالعودة إلى المثال 3.3، والبيانات المعطاة في جدول (2.6)، اعتبر النماذج التالية:

$$\ln (\text{hwage})_i = \beta_1 + \beta_2 \ln (\text{education})_i + \beta_3 (\ln \text{education})^2 + u_i$$

حيث ln = اللوغاريتم الطبيعي. كيف يمكنك تفسير هذا النموذج؟ قدر النموذج واحصل على الإحصاءات التقليدية وعلق على نتائجك.

(b) اعتبر الآن النموذج التالي:

$$\ln (\text{hwage}) = \beta_1 + \beta_2 \ln (\text{education}) + \beta_3 \ln (\text{educaton})^2 + u_i$$

إذا حاولت تقدير هذا النموذج، ما هي المشكلة (أو المشكلات) التي يمكن أن تتعرض لها؟ حاول تقدير هذا النموذج وحدد ما إذا كانت حزمة البرامج التي تستخدمها يمكنها القيام بتقدير هذا النموذج أم لا.

24.7 تجربة Monte Carlo. اعتبر النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

إذا علمت أن: $\beta_1 = 262$، $\beta_2 = -0.006$، $\beta_3 = -2.4$، $\sigma^2 = 42$ و $u_i \sim N(0, 42)$.
خلق 10 مجموعات من 64 مشاهدة من u_i باستخدام التوزيع المعتاد المعطى، واستخدم الـ 64 مشاهدة المعطاة في جدول (4.6). حيث Y=CM، X_2=PGNP، X_3=FLR لتخليق 10 مجموعات من القيم المقدرة للـ β (كل مجموعة سيكون لها ثلاثة معالم مقدرة). احصل على متوسط مقدرات معالمات الـ β، وادرس العلاقة بينها وبين القيم الحقيقية للـ β المعطاة أعلى. ما هو الاستنتاج العام الذي توصلت إليه؟

APPENDIX ملحق A7

1.A7 اشتقاق مقدرات OLS المعطاة في المعادلات (3.4.7) حتى (5.4.7)
DERIVATION OF OLS ESTIMATORS GIVEN IN EQUATIONS (7.4.3) TO (7.4.5)

فاضل المعادلة التالية:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})^2 \tag{2.4.7}$$

جزئيًا بالنسبة للثلاثة مجاهيل الموجودة في المعادلة، وساوى التفاضل في كل مرة بالصفر، ستحصل على:

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_1} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-1) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_2} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{2i}) = 0$$

$$\frac{\partial \sum \hat{u}_i^2}{\partial \hat{\beta}_3} = 2 \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i})(-X_{3i}) = 0$$

وبتبسيط هذه المعادلات، نحصل على المعادلة (3.4.7) حتى (5.4.7) وعمومًا لاحظ أن الثلاث معادلات السابقة يمكن أيضًا كتابتها كالتالي:

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = 0 \quad \text{(لماذا؟)}$$

$$\sum \hat{u}_i X_{3i} = 0$$

وذلك يوضح خواص مقدرات المربعات الصغرى، وهي أن مجموع البواقي يساوي الصفر وهذه البواقي غير مرتبطة مع المتغيرات المفسرة X_2 و X_3.

وبهذه المناسبة لاحظ أنه من أجل الحصول على مقدرات OLS في حالة نموذج انحدار خطي به k متغير (7.4.20) سنقوم بعمل نفس الخطوات، وبالتالي سنقوم أولاً بالتالي:

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \cdots - \hat{\beta}_k X_{ki})^2$$

وبالحصول على التفاضل الجزئي للمعادلة السابقة بالنسبة لكل مجهول من الـ k مجهول الموجودة في المعادلة، ومساواة ذلك التفاضل بالصفر، ثم إعادة ترتيب المعادلة نحصل على k من المعادلات الطبيعية في k مجهول كالتالي:

$$\sum Y_i = n\hat{\beta}_1 + \hat{\beta}_2 \sum X_{2i} + \hat{\beta}_3 \sum X_{3i} + \cdots + \hat{\beta}_k \sum X_{ki}$$

$$\sum Y_i X_{2i} = \hat{\beta}_1 \sum X_{2i} + \hat{\beta}_2 \sum X_{2i}^2 + \hat{\beta}_3 \sum X_{2i} X_{3i} + \cdots + \hat{\beta}_k \sum X_{2i} X_{ki}$$

$$\sum Y_i X_{3i} = \hat{\beta}_1 \sum X_{3i} + \hat{\beta}_2 \sum X_{2i} X_{3i} + \hat{\beta}_3 \sum X_{3i}^2 + \cdots + \hat{\beta}_k \sum X_{3i} X_{ki}$$

. . . .

$$\sum Y_i X_{ki} = \hat{\beta}_1 \sum X_{ki} + \hat{\beta}_2 \sum X_{2i} X_{ki} + \hat{\beta}_3 \sum X_{3i} X_{ki} + \cdots + \hat{\beta}_k \sum X_{ki}^2$$

وباستخدام الحروف الصغيرة للترميز، هذه المعادلات تصبح كالتالي:

$$\sum y_i x_{2i} = \hat{\beta}_2 \sum x_{2i}^2 + \hat{\beta}_3 \sum x_{2i} x_{3i} + \cdots + \hat{\beta}_k \sum x_{2i} x_{ki}$$

$$\sum y_i x_{3i} = \hat{\beta}_2 \sum x_{2i} x_{3i} + \hat{\beta}_3 \sum x_{3i}^2 + \cdots + \hat{\beta}_k \sum x_{3i} x_{ki}$$

. . . .

$$\sum y_i x_{ki} = \hat{\beta}_2 \sum x_{2i} x_{ki} + \hat{\beta}_3 \sum x_{3i} x_{ki} + \cdots + \hat{\beta}_k \sum x_{ki}^2$$

ويمكن ملاحظة أيضًا أن النموذج ذو k متغير تتحقق أيضًا فيه المعادلات التالية:

$$\sum \hat{u}_i = 0$$

$$\sum \hat{u}_i X_{2i} = \sum \hat{u}_i X_{3i} = \cdots = \sum \hat{u}_i X_{ki} = 0$$

2.A7 تساوي معاملات PGNP في (5.3.7) و (2.6.7)

EQUALITY BETWEEN THE COEFFICIENTS OF PGNP IN (7.3.5) AND (7.6.2)

دع $CM=Y$، $PGNP=X_2$ و $FLR=X_3$ وباستخدام التفاضلات السابقة، اكتب

$$y_i = b_{13} x_{3i} + \hat{u}_{1i} \quad (1)$$

$$x_{2i} = b_{23} x_{3i} + \hat{u}_{2i} \quad (2)$$

والآن قم بعمل انحدار لـ $\hat{u}_1$ على $\hat{u}_2$ فنحصل على

$$a_1 = \frac{\sum \hat{u}_{1i} \hat{u}_{2i}}{\hat{u}_{2i}^2} = -0.0056 \quad \text{(بالنسبة لمثالنا)} \quad (3)$$

لاحظ أنه بما أن $\hat{u}$'s هي البواقي، فإن متوسطتها يساوي العنصر باستخدام (1) و (2) يمكن كتابة (3) كالتالي:

$$a_1 = \frac{\sum (y_i - b_{13} x_{3i})(x_{2i} - b_{23} x_{3i})}{\sum (x_{2i} - b_{23} x_{3i})^2} \quad (4)$$

وبتطبيق نفس الفكرة يمكن ملاحظة أن:

$$b_{23} = \frac{\sum x_{2i} x_{3i}}{\sum x_{3i}^2} \quad (5)$$

و

$$b_{13} = \frac{\sum y_i x_{3i}}{\sum x_{3i}^2} \tag{6}$$

وبالتعويض عنها في (4)، نحصل على

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2} \tag{7.4.7}$$

$= -0.0056$ (بالنسبة لمثالنا الحالي) $=$

3.A7 اشتقاق المعادلة (19.4.7) :

DERIVOTION OF EQUATION (7.4.19)

تذكر أن :

$$\hat{u}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_{2i} - \hat{\beta}_3 X_{3i}$$

والذي يمكن أيضًا كتابته كالتالي :

$$\hat{u}_i = y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}$$

حيث الحروف الصغيرة، كالعادة، تعبر عن الانحرافات عن القيم المتوسطة، والآن

$$\begin{aligned}\sum \hat{u}_i^2 &= \sum (\hat{u}_i \hat{u}_i) \\ &= \sum \hat{u}_i (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i}) \\ &= \sum \hat{u}_i y_i\end{aligned}$$

حيث تم استخدام حقيقة أن (لماذا؟) $\sum \hat{u}_i x_{2i} = \sum \hat{u}_i x_{3i} = 0$

أيضًا :

$$\sum \hat{u}_i y_i = \sum y_i \hat{u}_i = \sum y_i (y_i - \hat{\beta}_2 x_{2i} - \hat{\beta}_3 x_{3i})$$

بمعنى أن :

$$\sum \hat{u}_i^2 = \sum y_i^2 - \hat{\beta}_2 \sum y_i x_{2i} - \hat{\beta}_3 \sum y_i x_{3i}$$

وهذه هي النتيجة المطلوبة .

4.A7 مقدرات الإمكان الأعظم لنموذج الانحدار المتعدد :

MAXIMUM LIKELIHOOD ESTIMATION OF THE MULTIPLE REGRESSION MODEL

لتوسيع دائرة تطبيق الأفكار المقدمة في الفصل (4)، ملحق A4، يمكننا كتابة لوغاريتم دالة الإمكان بالنسبة لنموذج انحدار خطي ذو k متغير (20.4.7) كالتالي :

$$\ln L = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum\frac{(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2}{\sigma^2}$$

وبالحصول علي التفاضل الجزيئي لهذه الدالة بالنسبة لكل من β_1، β_2، . . . ، β_k و σ^2 نحصل على $(k+1)$ من المعادلات كالتالي :

$$\frac{\partial \ln L}{\partial \beta_1} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-1) \qquad (1)$$

$$\frac{\partial \ln L}{\partial \beta_2} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{2i}) \qquad (2)$$

. . . .

$$\frac{\partial \ln L}{\partial \beta_k} = -\frac{1}{\sigma^2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})(-X_{ki}) \qquad (K)$$

$$\frac{\partial \ln L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \cdots - \beta_k X_{ki})^2 \qquad (K+1)$$

وبمساواة كل معادلة من هذه المعادلات السابقة بالصفر (شرط التنظيم من الدرجة الأولى) وباعتبار $\tilde{\sigma}^2$، $\tilde{\beta}_k$ ، . . . ، $\tilde{\beta}_2$، $\tilde{\beta}_1$ مقدرات ML وبعد إجراء بعض العمليات الجبرية البسيطة نحصل على التالي :

$$\sum Y_i = n\tilde{\beta}_1 + \tilde{\beta}_2\sum X_{2i} + \cdots + \tilde{\beta}_k\sum X_{ki}$$

$$\sum Y_i X_{2i} = \tilde{\beta}_1\sum X_{2i} + \tilde{\beta}_2\sum X_{2i}^2 + \cdots + \tilde{\beta}_k\sum X_{2i}X_{ki}$$

. . . .

$$\sum Y_i X_{ki} = \tilde{\beta}_1\sum X_{ki} + \tilde{\beta}_2\sum X_{2i}X_{ki} + \cdots + \tilde{\beta}_k\sum X_{ki}^2$$

وكما ترى، فتلك هي المعادلات الطبيعية لنظرية المربعات السابقة كما سبق ورأيناها في الملحق A7، فقرة 1.A7. وبالتالي فمقدرات ML للـ $\tilde{\beta}$'s هي نفس مقدرات الـ OLS للـ $\hat{\beta}$'s المعطاة سابقًا. ولكن كما سبق وذكرنا في الفصل (4)، ملحق A4 هذه التساوي ليس مجرد مصادفة.

بالتعويض عن مقدرات ML (=OLS) في الـ $(K+1)$ معادلة المعطاة سابقًا، وبعد إجراء بعض التبسيط، نحصل على مقدر ML لـ σ^2 كالتالي:

$$\tilde{\sigma}^2 = \frac{1}{n}\sum(Y_i - \tilde{\beta}_1 - \tilde{\beta}_2 X_{2i} - \cdots - \tilde{\beta}_k X_{ki})^2$$
$$= \frac{1}{n}\sum \hat{u}_i^2$$

كما سبق وذكرنا، هذا المقدر مختلف عن مقدر OLS لـ $\hat{\sigma}^2$، وهذا الأخير يساوي $\sum \hat{u}_i^2/(n-k)$. وبما أن الأخير يعتبر مقدرًا غير متحيز لـ σ^2، فإن ذلك يعني أن مقدار ML لـ $\tilde{\sigma}^2$ هو مقدر متحيز، ولكن يمكن إثبات أن $\tilde{\sigma}^2$ غير متحيز تقاربيًا أيضًا.

5.A7 مخرجات SAS لدالة إنتاج Cobb-Douglas (4.9.7):

SAS OUTPUT OF THE COBB-DOUGLAS PRODUCTION FUNCTION (7.9.4)

DEP VARIABLE: Y1

SOURCE	DF	SUM OF SQUARES	MEAN SQUARE	F VALUE	PROB > F
MODEL	2	0.538038	0.269019	48.069	0.0001
ERROR	12	0.067153	0.005596531		
C TOTAL	14	0.605196			

ROOT MSE	0.074810	R-SQUARE	0.8890
DEP MEAN	10.096535	ADJ R-SQ	0.8705
C.V.	0.7409469		

VARIABLE	DF	PARAMETER ESTIMATE	STANDARD ERROR	T FOR HO: PARAMETER = 0	PROB > |T|
INTERCEP	1	−3.338455	2.449508	−1.363	0.1979
Y2	1	1.498767	0.539803	2.777	0.0168
Y3	1	0.489858	0.102043	4.800	0.0004

COVARIANCE OF ESTIMATES

COVB	INTERCEP	Y2	Y3
INTERCEP	6.000091	−1.26056	0.1121951
Y2	−1.26056	0.2913868	−0.0384272
Y3	0.01121951	−0.0384272	0.01041288

Y	X2	X3	Y1	Y2	Y3	Y1HAT	Y1RESID
16607.7	275.5	17803.7	9.7176	5.61859	9.7872	9.8768	−0.15920
17511.3	274.4	18096.8	9.7706	5.61459	9.8035	9.8788	−0.10822
20171.2	269.7	18271.8	9.9120	5.59731	9.8131	9.8576	0.05437
20932.9	267.0	19167.3	9.9491	5.58725	9.8610	9.8660	0.08307
20406.0	267.8	19647.6	9.9236	5.59024	9.8857	9.8826	0.04097

20831.6	275.0	20803.5	9.9442	5.61677	9.9429	9.9504	−0.00615
24806.3	283.0	22076.6	10.1189	5.64545	10.0023	10.0225	0.09640
26465.8	300.7	23445.2	10.1836	5.70611	10.0624	10.1428	0.04077
27403.0	307.5	24939.0	10.2184	5.72848	10.1242	10.2066	0.01180
28628.7	303.7	26713.7	10.2622	5.71604	10.1929	10.2217	0.04051
29904.5	304.7	29957.8	10.3058	5.71933	10.3075	10.2827	0.02304
27508.2	298.6	31585.9	10.2222	5.69910	10.3605	10.2783	−0.05610
29035.5	295.5	33474.5	10.2763	5.68867	10.4185	10.2911	−0.01487
29281.5	299.0	34821.8	10.2847	5.70044	10.4580	10.3281	−0.04341
31535.8	288.1	41794.3	10.3589	5.66331	10.6405	10.3619	−0.00299

COLLINEARITY DIAGNOSTICS **VARIANCE PROPORTIONS**

NUMBER	EIGENVALUE	CONDITION INDEX	PORTION INTERCEP	PORTION Y2	PORTION Y3
1	3.000	1.000	0.0000	0.0000	0.0000
2	.000375451	89.383	0.0491	0.0069	0.5959
3	.000024219	351.925	0.9509	0.0031	0.4040

DURBIN-WATSON *d*	0.891
1ST ORDER AUTOCORRELATION	0.366

مـلاحظات: ln Y=Y1 ، ln Y2=X2 ، ln X3=Y3 . القـيـم المعطاة تحت العنوان PROB>|T| تمثل الـ P-values . انظر الفصل (15) لمزيد من التفاصيل عن كيفية توصيف مشكلة الارتباط الخطي.

الفصل الثامن

تحليل الانحدار المتعدد.. مشكلة الاستدلال

MULTIPLE REGRESSION ANALYSIS: THE PROBLEM OF INFERENCE

هذا الفصل يعتبر تابعًا للفصل (5)، حيث سيتم توسيع دائرة تطبيق أفكار التقدير بفترة، واختيارات الفروض التي تم تناولتها في الفصل (5)، لتشمل النماذج التي تحتوي على ثلاثة متغيرات أو أكثر. على الرغم من أنه في كثير من الأحيان، يمكن تطبيق نفس المفاهيم التي تنا ولها الفصل (5)، على حالة الانحدار المتعدد تطبيق مباشر، إلا أن هناك بعض الخواص القليلة الإضافية الخاصة بمثل هذه النماذج المتعددة، والتي ستحصل على الاهتمام الأكبر في هذا الفصل.

1.8 مرة أخرى فرض التوزيع المعتاد:
THE NORMALITY ASSUMPTION ONCE AGAIN

نحن نعرف الآن أنه إذا كان هدفنا الوحيد هو التقدير بنقطة لمعالم نموذج الانحدار، فإن طريقة المربعات الصغرى العادية (OLS) والتي لا تضع أي فروض على التوزيع الاحتمالي للخطأ العشوائي u_i هي مقدرات كافية للتقدير. أما إذا كان هدفنا هو التقدير والاستدلال معاً، فإننا نحتاج إلى فرض خاص بـ u_i حول التوزيع الاحتمالي له. وقد تناولنا ذلك من قبل في الفصلين (4 و5).

ولأسباب سبق ذكرها، افترضنا أن u_i يتبع التوزيع المعتاد بتوقع يساوي الصفر، وتباين ثابت σ^2. سنستكمل استخدام مثل هذا الفرض في حالة نماذج الانحدار المتعدد. مع فرض التوزيع المعتاد. ووفقاً لما تناولناه في الفصلين (4 و 7)، سنجد أن مقدرات OLS لمعاملات الانحدار الجزئية هي أفضل المقدرات الخطية

غير المتحيزة (BLUE)[1]. والأكثر من ذلك، فإن مقدرات $\hat{\beta}_2$ ، $\hat{\beta}_3$ و $\hat{\beta}_1$ هي نفسها تتبع التوزيع المعتاد بتوقع يساوي المعلمة الحقيقية $\hat{\beta}_2$، $\hat{\beta}_3$، $\hat{\beta}_1$ وتباينها معطى في الفصل (7).

بالإضافة إلى $(n-3)\hat{\sigma}^2/\sigma^2$ والتي تتبع توزيع X^2 بدرجات حرية تساوي $n-3$ ومقدرات OLS الثلاثة مستقلة عن $\hat{\sigma}^2$. إثبات ذلك يمكن عمله باتباع نفس خطوات حالة وجود متغيرين اثنين فقط، والتي تم مناقشتها في ملحق 3. وكنتيجة لذلك، واتباعاً للفصل (5)، يمكن إثبات أن بعد استبدال قيمة σ^2 بمقدرها غير المتحيز $\hat{\sigma}^2$ عند حساب الأخطاء القياسية، فإن كلاً من المتغيرات التالية:

$$t = \frac{\hat{\beta}_1 - \beta_1}{\text{se}(\hat{\beta}_1)} \tag{1.1.8}$$

$$t = \frac{\hat{\beta}_2 - \beta_2}{\text{se}(\hat{\beta}_2)} \tag{2.1.8}$$

$$t = \frac{\hat{\beta}_3 - \beta_3}{\text{se}(\hat{\beta}_3)} \tag{3.1.8}$$

يتبع توزيع t بدرجات حرية تساوي $n-3$.

لاحظ أن درجات الحرية الآن تساوي $n-3$، حيث إنه عند حساب $\sum \hat{u}_i^2$، وبالتالي $\hat{\sigma}^2$ نحتاج أولاً إلى تقدير معاملات الانحدار الجزيئية الثلاثة، مما يجعلنا نضع ثلاثة قيود على مجموع مربعات البواقي RSS (وباتباع نفس المنطق، فإنه في حالة وجود أربعة متغيرات، تكون درجات الحرية مساوية $n-4$ وهكذا). وبالتالي، فإن توزيع t يمكن استخدامه للحصول على فترات ثقة وإجراء اختبارات فروض حول معالم الانحدار الجزيئية الحقيقية الخاصة بالمجتمع.

وبالمثل، فإن توزيع X^2 يمكن استخدامه لعمل اختبارات فروض حول قيمة σ^2 الحقيقية. لتوضيح الخطوات الفعلية لذلك، دعنا نستعرض المثال التوضيحي التالي.

---

(1) مع افتراض اتباع الخطأ للتوزيع المعتاد، فإن مقدرات OLS لكل من $\hat{\beta}_2$ ، $\hat{\beta}_3$ و $\hat{\beta}_1$ هي مقدرات ذات تباين أقل داخل فئة المقدرات غير المتحيزة، سواء خطية أو غير خطية. باختصار، فإن هذه المقدرات يطلق عليها BUE (أفضل المقدرات غير المتحيزة). انظر
C. R. Rao, Linear Statistical Inference and Its Applications, John Wiley & Sons, New York, 1965, p. 258.

2.8 مثال 1.8 : مرة أخرى مثال وفيات الأطفال:
CHILD MORTALITY EXAMPLE REVISITED

في الفصل (7)، قمنا بعمل انحدار لوفيات الأطفال (CM) على GNP الفردي (PGNP)، ومعدل حرية المرأة (FLR) على عينة من 64 بلداً. نتائج الانحدار المعطاة في (2.6.7) معطاة مع بعض الإضافات كالتالي :

$$\begin{aligned}
\widehat{\text{CM}}_i &= 263.6416 \quad - \quad 0.0056\ \text{PGNP}_i - \quad 2.2316\ \text{FLR}_i \\
\text{se} &= (11.5932) \qquad (0.0019) \qquad (0.2099) \\
t &= (22.7411) \qquad (-2.8187) \qquad (-10.6293) \qquad (1.2.8) \\
p \text{ value} &= (0.0000)^* \qquad (0.0065) \qquad (0.0000)^* \\
& R^2 = 0.7077 \qquad \bar{R}^2 = 0.6981
\end{aligned}$$

حيث : * تعني قيمة صغيرة للغاية.

في المعادلة (1.2.8) تتبعنا نفس الشكل الذي قدمناه سابقاً في المعادلة (1.11.5)، حيث إن الأرقام في المجموعة الأولى من الأقواس، هي تقديرات الأخطاء القياسية، أما الموجودة في المجموعة الثانية فهي قيم t للفرض الهدمي الخاص بأن معلمة المجتمع الحقيقية تساوي الصفر، والأرقام الموجودة بين الأقواس في المجموعة الثالثة تمثل قيم p-values. ومعطى أيضاً R^2 وR^2 المصححة. وقد فسرنا هذا الانحدار من قبل في مثال 1.7 .

ماذا عن المعنوية الإحصائية لهذه النتائج؟ فمثلاً دعنا نعتبر معامل PGNP المساوي لـ 0.0056-. هل هذا المعامل معنوي إحصائياً؟ أي يختلف إحصائياً عن الصفر؟ هل كل من المعاملين له معنوية إحصائية؟ للإجابة على هذا السؤال وتساؤلات أخرى مرتبطة به، دعنا أولاً نعتبر أنواع اختبارات الفروض المختلفة التي قد يحتاج إليها الفرد في إطار نماذج الانحدار المتعددة.

3.8 اختبارات الفروض في إطار الانحدار المتعدد:
HYPOTHESIS TESTING IN MULTIPLE REGRESSION :

تعليقات عامة : General comments

بمجرد أن خرجنا خارج دائرة العالم البسيط لنموذج الانحدار الخطي ثنائي المتغيرات، يصبح لاختبارات الفروض أشكال عديدة مهمة، كالتالي :

1 - اختبارات فروض حول معامل انحدار جزيئي واحد (فقرة 4.8).

2 - اختبار معنوية نموذج الانحدار المتعدد المقدر ككل، بمعنى معرفة ما إذا كانت معاملات الميل الجزيئية تساوي الصفر آنياً أم لا. (فقرة 5.8)

3 - اختبار معاملين أو أكثر متساويان أم لا. (فقرة 6.8)

4 - اختبار ما إذا كان معامل انحدار جزيئي ما يحقق شروطًا معينة. (فقرة 7.8)

5 - اختبار سكون معادلة الانحدار المقدرة مع مرور الزمن أو وفقاً لوحدات مستعرضة مختلفة. (فقرة 8.8)

6 - اختبار الشكل الدالي لنموذج الانحدار. (فقرة 9.8)

ونظراً لأن واحداً أو أكثر من هذه الاختبارات يظهر كثيراً في التحليل العملي، فقد خصصنا فقرة لكل نوع.

4.8 اختبارات الفروض حول معاملات الانحدار الفردية: HYPOTHESIS TESTING ABOUT INDIVIDUAL REGRESSION COEFFICIENTS

إذا فرضنا أن $u_i \sim N(0, \sigma^2)$، فإنه كما سبق وذكرنا في الفقرة 1.8 يمكننا استخدام اختبار t لاختبار أي فرض الخاص بأن أي معامل انحدار جزيئي فردي. لشرح آلية ذلك، اعتبر نموذج وفيات الأطفال (1.2.8) دعنا نفترض أن

$$H_0: \beta_2 = 0 \quad \text{and} \quad H_1: B_2 \neq 0$$

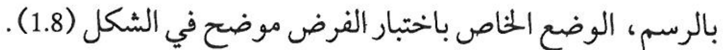
الفرض الهدمي ينص على أنه، بافتراض ثبات X_3 (معدل حرية المرأة)، فإن X_2 (PGNP) ليس له أي تأثير (خطي) على Y (وفيات الأطفال)[2]. لاختبار الفرض العدمي، نستخدم اختبار t المحسوبة تزيد عن قيمة t الجدولية (الحرجة) عند مستوى المعنوية المحدد، فإنه يمكن رفض الفرض العدمي. وبخلاف ذلك يتم عدم رفض الفرض العدمي. بالنسبة لمثالنا التوضيحي باستخدام (2.1.8) وملاحظة أن $\beta_2 = 0$ تحت صحة الفرض العدمي، فإننا نحصل على:

$$t = \frac{-0.0056}{0.0020} = -2.8187 \qquad (1.4.8)$$

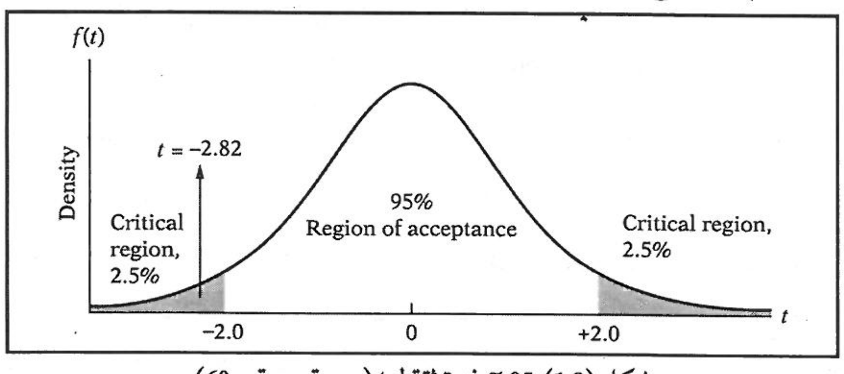
كما يتضح من المعادلة (1.2.8)

لاحظ أن لدينا 64 مشاهدة. وبالتالي فإن درجات الحرية في هذا المثال تساوي 61 (لماذا؟). إذا عدنا إلى جدول t المعطى في ملحق D. ليس لدينا بيانات عن درجات حرية تساوي 61. الأقرب عند 60، إذا استخدمناها وافترضنا قيمة α عند 0.05 (أي احتمال الوقوع في الخطأ من النوع الأول) فإن قيمة t الجدولية تكون 2 بالنسبة لاختبار ذي طرفين (انظر على $t_{\frac{\alpha}{2}}$ باستخدام 60 درجة حرية) أو 1.671 بالنسبة لاختبار ذي طرف واحد (انظر على t_{α} باستخدم 60 درجة حرية).

بالنسبة لمثالنا، الفرض البديل له طرفان. وبالتالي سنستخدم قيمة t ذات الطرفين. بما أن قيمة t، المحسوبة والمساوية لـ 2.8187 (القيمة المطلقة) تزيد عن قيمة t الجدولية والمساوية لـ 2، فإننا نرفض الفرض العدمي القائل بأن PGNP ليس له أي تأثير على وفيات الأطفال. لتوضيح ذلك، بافتراض ثبات معدل حرية المرأة، فإن GNP له تأثير معنوي (عكسي) على وفيات الأطفال، وذلك متوقع مسبقاً قبل إجراء أي تحليل.



بالرسم، الوضع الخاص باختبار الفرض موضح في الشكل (1.8).



شكل (1.8) 95% فترة ثقة لـ t (درجة حرية = 60)

عملياً، ليس بالضرورة فرض قيمة محددة لـ α لإجراء اختبارات الفروض. فممكن ببساطة استخدام قيمة p-value المعطاة في (2.2.8) والتي تساوي 0.0065.

(2) في معظم الأبحاث التطبيقية يتم عرض الفرض العدمي في تلك الصورة، بمعنى اتخاذ أعلى شيء، بمعنى أنه لا توجد علاقة بين المتغير التابع والمتغير المفسر محل الدراسة. الفكرة هنا هي إيجاد ما إذا كانت هناك علاقة أصلاً بين المتغيرين نستطيع أن نعتمد عليها أم لا.

وتفسير الـ p-value (أي مستوى المعنوية التام) على أنها احتمال الحصول على قيمة t (كقيمة مطلقة) مساوية لـ 2.8187 أو أكثر، بافتراض صحة الفرض العدمي، تساوي 0.0065 أو 0.65%، وتلك القيمة تعتبر قيمة صغيرة وأصغر كثيراً من قيمة α المساوية لـ 5%.

هذا المثال يعطينا فرصة لاختيار ما إذا كان بصدد اختبار ذي طرف واحد، أو اختبار ذي طرفين. بما أنه متوقع مسبقاً أن هناك علاقة عكسية بين وفيات الأطفال والـ GNP الفردي (لماذا؟)، فإنه يجب اختيار اختبار ذي طرف واحد. أي أن الفرض العدمي والبديل سيكونان كالتالي :

$$\beta_0 : \beta_2 < 0 \quad \text{and} \quad H_1 : \beta_2 \geq 0$$

وكما يعلم القارئ، يمكننا رفض الفرض العدمي على أساس اختبار ذي طرف واحد في هذا المثال.

في الفصل (5)، رأينا العلاقة بين اختبارات الفروض من جهة، وفترات الثقة من جهة أخرى. ففي مثالنا الحالي، 95% فترة ثقة لـ β_2 هي :

$$\hat{\beta}_2 - t_{\frac{a}{2}} \operatorname{se}(\hat{\beta}_2) \leq \beta_2 \leq \hat{\beta}_2 + t_{\frac{a}{2}} \operatorname{se}(\hat{\beta}_2)$$

والتي تساوي في مثالنا الحالي

$$-0.0056 - 2\ (0.0020) \leq \beta_2 \leq -0.0056 + 2\ (0.0020)$$

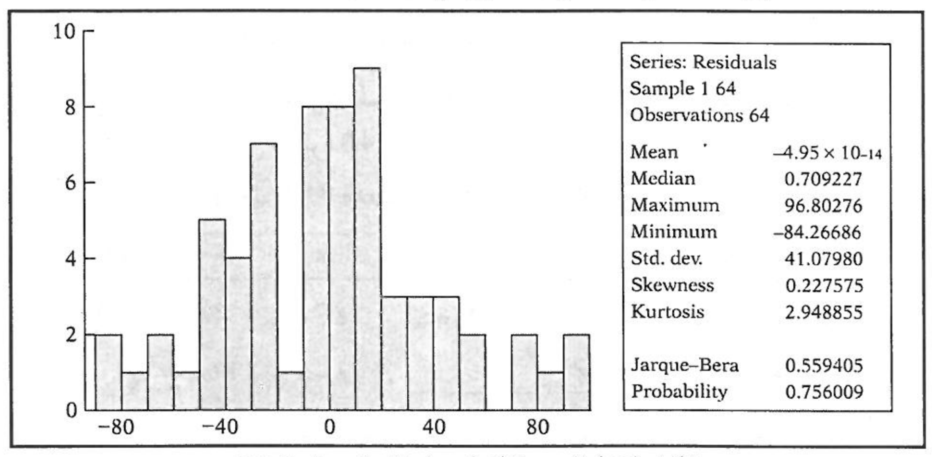
وهذا يساوي

$$-0.0096 \leq \beta_2 \leq -0.0016 \qquad \textbf{(2.4.8)}$$

أي أن الفترة 0.0096- إلى 0.0016- تشتمل على القيمة الحقيقية لمعامل β_2 بدرجة ثقة 95%. أي أنه إذا اخترنا 100 عينة لها الحجم 64 وكون 100 فترة ثقة مثل الموجودة في (2.4.8)، نتوقع أن 95 منهم سيهتدون على معلمة المجتمع الحقيقية β_2. بما أن الفترة (2.4.8) لا تشتمل على قيمة الصفر الخاصة بالفرض العدمي، فإننا نرفض الفرض العدمي القائل بأن قيمة β_2 الحقيقية مساوية للصفر بدرجة ثقة 95%.

وبالتالي، سواء استخدمنا اختبار المعنوية t كما في (1.4.8) أو التقدير بفترة ثقة، كما في (2.4.8)، سنصل إلى نفس الاستنتاج، عموماً هذه النتيجة غير مستغربة مع اعتبار العلاقة بين فترات الثقة واختبارات الفروض.

باتباع نفس الأسلوب المذكور أعلاه، يمكننا عمل اختبارات فروض حول المعلمات الأخرى الموجودة في نموذج انحدار وفيات الأطفال. البيانات الضرورية معطاة بالفعل في المعادلة (1.2.8). فعلى سبيل المثال، إذا أردنا اختبار أنه لا يوجد تأثير لمعدل حرية المرأة على وفيات الأطفال بافتراض ثبات BGNP - يمكننا بثقة رفض هذه الفرضية، حيث إنه تحت صحة الفرض العدمي قيمة p-value لقيمة t المطلقة المساوية لـ 10.6 أو أكثر هي مساوية للصفر. قبل الانتقال لنقطة أخرى، تذكر أن خطوات اختبار t تعتمد على افتراض أن مقدار الخطأ u_i يتبع التوزيع المعتاد. وعلى الرغم من أننا لا نستطيع مباشرة مشاهدة u_i، يمكننا مشاهدة $\hat{u}_i$، أي البواقي. في انحدار الوفيات، المدرج التكراري للبواقي موضح في الشكل (2.8).



شكل (2.8) المدرج التكراري لبواقي انحدار (1.2.8)

من المدرج التكراري، نلاحظ أن البواقي تتبع التوزيع المعتاد. يمكننا أيضاً حساب اختبار Jarque- Bera JB لاختبار فرض التوزيع المعتاد، كما هو موضح في المعادلة (1.12.5).

في مثالنا الحالي، قيمة JB تساوي 0.5594 ولها قيمة p-value تساوي 0.76. [3] وبالتالي، يبدو أن مقدار الخطأ في مثالنا الحالي يتبع التوزيع المعتاد. وبالطبع ضع في الاعتبار أن اختبار JB هو اختبار للعينات كبيرة الحجم، وحجم العينة في مثالنا هذا هو 64 مشاهدة، والذي قد لا يكون بالضرورة حجم عينة كبير.

(3) في مثالنا الحالي ، قيمة معامل الالتواء هي 0.2276، وقيم معامل التفلطح هي 2.9488 تذكر أن قيم معاملات الالتواء والتفلطح للمتغيرات التي تتبع التوزيع المعتاد هي 0 ، 3 بالترتيب .

5.8 اختبار المعنوية الكلية لانحدار العينة :
TESTING THE OVERALL SIGNIFICANCE OF THE SAMPLE REGRESSION

في الفقرة السابقة، كنا مهتمين باختيار معنوية معاملات الانحدار الجزيئية المقدرة فردياً، أي اختبار فروض منفصلة لكل معامل انحدار جزيئي حقيقي للمجتمع إذا ما كان يساوي الصفر أم لا. ولكن الآن دعنا نعتبر الفرض التالي :

$$H_0: \beta_2 = \beta_3 = 0 \qquad (1.5.8)$$

هذا الفرض العدمي وهو فرض مشترك لكل من β_2 وβ_3 حول ما إذا كانا يساويان الصفر آنياً. اختبار مثل هذا الفرض يسمى اختبار معنوية العلاقة ككل لمعادلة الانحدار المقدرة، أي اختبار ما إذا كانت Y لها علاقة خطية مع كل من X_2 و X_3 أم لا.

هل يمكن للفرض المشترك الموجود في (1.5.8) اختباره من خلال اختبار معنوية $\hat{\beta}_2$ ، $\hat{\beta}_3$ منفردين كما في الفقرة 4.8؟ الإجابة هي لا، والأسباب كالتالي :

جدول (1.8) جدول ANOVA لانحدار ثلاثي المتغيرات

Source of variation	SS	df	MSS
Due to regression (ESS)	$\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}}{2}$
Due to residual (RSS)	$\sum \hat{u}_i^2$	$n-3$	$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-3}$
Total	$\sum y_i^2$	$n-1$	

تذكر المعادلة التالية :

$$\sum y_i^2 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i} + \sum \hat{u}_i^2 \qquad (2.5.8)$$

$$\text{TSS} = \quad \text{ESS} \quad + \text{RSS}$$

TSS له كالعادة، $n-1$ درجة حرية وRss له $n-3$ درجة حرية لأسباب سبق وذكرناها تفصيليًا. ESS له 2 درجة حرية، حيث إنه دالة في $\hat{\beta}_2$ و$\hat{\beta}_3$. وبالتالي باتباع أسلوب الـ ANOVA الذي ناقشناه في 9.5، يمكننا تكوين جدول (1.8)، الآن يمكننا إثبات [6] أنه بافتراض اتباع الـ u_i للتوزيع المعتاد وبافتراض صحة الفرض العدمي الخاص بـ $\beta_2 = \beta_3 = 0$، فإن المتغير

(6) انظر K.A. Brownlee, Statisitical Theory and Methodology in Science and Engineering, John Wiley & Sons, New York, 1960, pp. 278–280.

$$F = \frac{\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)/2}{\sum \hat{u}_i^2/(n-3)} = \frac{\text{ESS/df}}{\text{RSS/df}} \quad \textbf{(3.5.8)}$$

له توزيع F بدرجات حرية 2، $n-3$.

ما الذي يمكننا الاستفادة به من نسبة F السابقة؟ يمكننا إثبات أن[7] تحت صحة الفرض الخاص بأن $u_i \sim N(0, \sigma^2)$ فإن

$$E\frac{\sum \hat{u}_i^2}{n-3} = E(\hat{\sigma}^2) = \sigma^2 \quad \textbf{(4.5.8)}$$

مع الفرض الإضافي الخاص بأن $\beta_2 = \beta_3 = 0$، يمكننا إثبات أن :

$$\frac{E\left(\hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}\right)}{2} = \sigma^2 \quad \textbf{(5.5.8)}$$

وبالتالي إذا كان الفرض العدمي صحيحًا، فإن كلاً من (4.5.8) و(5.5.8) تعطيان تقديراً متساويًا للقيمة الحقيقية σ^2. هذه العبارة يجب ألا تكون مستغربة، حيث إن هناك علاقة مبدئية بين Y و X_2 و X_3 . فمصدر التباين الوحيد لـ Y سيرجع إلى المقدار العشوائي الممثل في u_i. إذا كان الفرض العدمي غير صحيح، أي أن X_2 و X_3 لهما تأثير على Y، فإن التساوي بين (4.5.8) و(5.5.8) لا يتحقق. في هذه الحالة يكون ESS أكبر نسبياً من RSS مع وضع درجات الحرية في الاعتبار. وبالتالي قيمة F الموجودة في (3.5.8) تمثل اختباراً للفرض العدمي الخاص بمعاملات الميل الحقيقية وتساويها آنياً بالصفر. إذا كانت قيمة F المحسوبة من (3.5.8) تزيد عن قيمة F الجدولية من جدول F بعد مستوى المعنوية α، فإننا سنرفض الفرض العدمي، وبخلاف ذلك لا نرفض الفرض العدمي. وبطريقة بديلة إذا كانت قيمة p-value لـ F المحسوبة صغيرة بشكل كاف، فإننا نرفض H_0.

جدول (2.8) يلخص اختبار F. وبالعودة إلى مثالنا التوضيحي، نحصل على جدول الـ ANOVA كما هو موضح في جدول (3.8).

باستخدام (3.5.8) نحصل على

$$F = \frac{128{,}681.2}{1742.88} = 73.8325 \quad \textbf{(6.5.8)}$$

قيمة p-value الخاصة بقيمة F المساوية لـ 73.8325 أو أكبر تساوي تقريباً الصفر، مما يجعلنا نرفض الفرض الخاص بأن PGNP وFLR معاً ليس لهما تأثير على وفيات الأطفال. إذا كنا سنستخدم مستوى معنوية 5%، فإن قيمة F الجدولية بدرجات

(7) المرجع السابق .

حرية 2 في البسط و60 درجة في المقام (أو فعلياً 61) هي 3.15 أو حوالي 4.98 إذا كنت تستخدم 1% مستوى معنوية. يتضح من ذلك أن F المحسوبة حوالي 74 مرة أكثر من أي من هذه القيم الجدولية لـ F.

جدول (2.8) تلخيص إحصاء F

Null hypothesis H_0	Alternative hypothesis H_1	Critical region Reject H_0 if
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 > \sigma_2^2$	$\frac{S_1^2}{S_2^2} > F_{\alpha,ndf,ddf}$
$\sigma_1^2 = \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$	$\frac{S_1^2}{S_2^2} > F_{\alpha/2,ndf,ddf}$ or $< F_{(1-\alpha/2),ndf,ddf}$

لاحظ أن :

1 - σ_1^2 و σ_2^2 هما تباين المجتمعتين.

2 - S_1^2 و S_2^2 هما تباين العينتين.

3 - ndf و ddf ترمزان إلى درجات حرية البسط والمقام بالترتيب.

4 - في حساب النسبة F، ضع قيمة S^2 الكبيرة في البسط.

5 - قيم F الجدولية موجودة في العمود الأخير. الترميز الأول للـ F هو مستوى المعنوية، والثاني هو درجات حرية البسط والمقام.

6 - لاحظ أن :

$$F_{(1-\alpha/2),\, ndf,\, ddf} = 1/F_{\alpha/2,\, ddf,\, ndf.}$$

جدول (3.8) . جدول ANOVA لمثال وفيات الأطفال

Source of variation	SS	df	MSS
Due to regression	257,362.4	2	128,681.2
Due to residuals	106,315.6	61	1742.88
Total	363,678	63	

اختبار معنوية العلاقة ككل في الانحدار المتعدد: اختبار F

قاعدة اتخاذ القرار في اختبارات الفروض Decision Rule

اعتبر نموذج الانحدار الذي يشتمل على k متغير كالتالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

لاختبار الفرض :

$$H_0 : \beta_2 = \beta_3 = \ldots = \beta_k = 0$$

(أي أن : معاملات الميل تساوي الصفر آنياً) ضد الفرض :

ليست جميع المعاملات مساوية للصفر آنياً : H_1

احسب :

$$F = \frac{\text{ESS/df}}{\text{RSS/df}} = \frac{\text{ESS}/(k-1)}{\text{RSS}/(n-k)} \qquad (7.5.8)$$

إذا كانت $F > F_\alpha(k-1, n-k)$، ارفض H_0، بخلاف ذلك لا ترفض H_0، حيث $F_\alpha(k-1, n-k)$ هي قيمة F الجدولية عند مستوى المعنوية α و$(k-1)$ درجات حرية البسط و$(n-k)$ درجات حرية المقام. بطريقة بديلة، إذا كانت قيمة p-value لـ F التي نحصل عليها من (7.5.8) صغيرة بشكل كاف، فإنه يمكنك رفض H_0.

ولا نحتاج إلى القول بأنه في حالة وجود ثلاثة متغيرات (Y, X_1, X_2) تكون k مساوية 3، في حالة أربعة متغيرات تكون k مساوية 4 وهكذا.

وعموماً لاحظ أنه معظم برامج الحاسب الخاصة بتحليل الانحدار تحسب قيمة F (معطاة في جدول تحليل التباين) مع مخرجات تحليل الانحدار العادية مثل مقدرات معاملات الانحدار، أخطاؤهم القياسية، قيم t وهكذا. الفرض العدمي عند حساب t يكون عادة يفترض أن $\beta_i = 0$.

اختبارات الفروض الفردية والكلية : Individual versus Joint Testing of Hypotheses

في فقرة 4.8 ناقشنا اختبارات الفروض الخاصة بمعاملات الانحدار الفردية، أي اختبار كل معامل على حدة. أما في الفقرة 5.8 ناقشنا اختبار معنوية العلاقة ككل أو المعنوية المشتركة للانحدار المقدر (أي معاملات الميل وتساويها مع الصفر آنياً). نكرر مرة أخرى، أن هذه الاختبارات مختلفة. أي أنه باستخدام اختبار t أو فترة الثقة كما في (الفقرة 4.8) من الممكن أن تقبل الفرض الخاص بمعلمة ميل ما، β_k وتساويها مع الصفر، ولكنك ترفض الفرض المشترك أن كل معاملات الميل تساوي الصفر.

الدرس الذي يمكن أن نتعلمه من ذلك هو أن «النتيجة» المشتركة لفترات الثقة الفردية ليست بديلاً عن منطقة الثقة المشتركة [المنصوص عليها في اختبار F] لتكوين اختبارات فروض مشتركة ونتائج شاملة[8].

(8) Fomby et al., op. cit., p. 42.

علاقة مهمة بين R^2 و F : An Important Relotionship

هناك علاقة وثيقة بين معامل التحديد R^2 واختبار F المستخدم في تحليل التباين. بافتراض أن مقدار الخطأ u_i يتبع التوزيع الطبيعي وتحت صحة الفرض العدمي $\beta_2 = \beta_3 = 0$ رأينا أن :

$$F = \frac{\text{ESS}/2}{\text{RSS}/(n-3)} \qquad \textbf{(8.5.8)}$$

هذا الإحصاء يتبع توزيع F بدرجات حرية 2 و $n-3$.

وبشكل أكثر تعميمًا، في حالة وجود k متغير (شامل الجزء الثابت)، إذا افترضنا أن الأخطاء تتبع التوزيع الطبيعي وتحت صحة الفرض العدمي

$$H_0:\ \beta_2 = \beta_3 = \ldots = \beta_k = 0 \qquad \textbf{(9.5.8)}$$

وبالتالي نجد أن :

$$F = \frac{\text{ESS}/(k-1)}{\text{RSS}/(n-k)} \qquad \textbf{(7.5.8) = (10.5.8)}$$

هذا الإحـصـاء يتـبـع توزيع F بدرجـات حـرية $k-1$ و $n-k$ (لاحظ أن : عـدد المعلمـات الكلي المراد تقـديره هو k، وهذا يشـمل مقـدار الجزء الثـابت) دعنا نقـوم ببعض العمليات الجبرية على (10.5.8) كالتالي :

$$\begin{aligned} F &= \frac{n-k}{k-1}\frac{\text{ESS}}{\text{RSS}} \\ &= \frac{n-k}{k-1}\frac{\text{ESS}}{\text{TSS}-\text{ESS}} \\ &= \frac{n-k}{k-1}\frac{\text{ESS/TSS}}{1-(\text{ESS/TSS})} \\ &= \frac{n-k}{k-1}\frac{R^2}{1-R^2} \\ &= \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \end{aligned} \qquad \textbf{(11.5.8)}$$

حيث تم استخدم تعريف R^2 على أنه يساوي ESS/TSS. المعادلة (11.5.8) توضح العلاقة بين F و R^2. فهاتان القيمتان تتغيران معاً بشكل مباشر. فعندما تكون $R^2 = 0$ فإن F أيضاً تساوي الصفر. وكلما زادت R^2 تزداد قيمة F، وهذا يعتبر مؤشراً لمعنوية معادلة الانحدار المقدرة ككل. وبشكل خاص عندما تكون $R^2 = 1$ فإن F تصل إلى

ما لا نهاية. وبالتالي، فإن اختبار F، والذي يعتبر مقياسًا لمعنوية معادلة الانحدار المقدرة ككل، يعتبر أيضاً اختباراً لمعنوية R^2. بمعنى آخر، اختبار الفرض العدمي (9.5.8) يساوي اختبار الفرض العدمي الخاص (بالمجتمع) والقائل بأن R^2 تساوي الصفر.

جدول (4.8) جدول ANOVA في صورة R^2

Source of variation	SS	df	MSS*
Due to regression	$R^2(\sum y_i^2)$	2	$R^2(\sum y_i^2)/2$
Due to residuals	$(1-R^2)(\sum y_i^2)$	$n-3$	$(1-R^2)(\sum y_i^2)/(n-3)$
Total	$\sum y_i^2$	$n-1$	

* لاحظ أن: في حساب قيمة F لا نحتاج لضرب كل من R^2 و$(1-R^2)$ بالقيمة Σy_i^2 لأنها تختفي كما هو موضح في (12.5.8)

في حالة وجود ثلاثة متغيرات (11.5.8) تصبح كالتالي :

$$F = \frac{R^2/2}{(1-R^2)/(n-3)} \qquad (12.5.8)$$

وباستخدام العلاقة الوثيقة في F، R^2 يمكن لجدول ANOVA (1.8) أن يوضح في شكل جدول (4.8).

في مثالنا التوضيحي، باستخدام (12.5.8) نحصل على :

$$F = \frac{0.7077/2}{(1-0.7077)/61} = 73.8726$$

وهذه القيمة قريبة مما تم الحصول عليه من قبل، باستثناء قيمة الخطأ المقرب.

من مميزات استخدام اختبار F في صورة R^2 سهولة الحساب، فكل ما يحتاج إليه الفرد هو معرفة قيمة R^2. وبالتالي فاختبار المعنوية الكلي F المعطي في (7.5.8) يمكن أن نحصل عليه في صورة R^2 كما هو موضح في جدول (4.8).

اختبار المعنوية الكلية للانحدار المتعدد في صورة R^2 :

Testing the Overall significance of a multiple Regression

قاعدة اتخاذ القرار : Decision Rule

اختبار المعنوية الكلية للانحدار في صورة R^2 يعتبر أسلوبًا بديلاً ومساويًا للاختبار الموجود في (7.5.8)

افترض أن لديك نموذج انحدار به k متغير كالتالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

لاختبار الفرض التالي :

$$H_0 : \beta_2 = \beta_3 = ... = \beta_k = 0$$

ضد

H_1 : ليست كل المعاملات السابقة تساوي الصفر آنياً

احسب

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} \qquad \textbf{(17.5.8)}$$

إذا كانت $F > F_{\alpha(k-1,\ n-k)}$ ارفض H_0، وبخلاف ذلك يمكن أن تقبل H_0، حيث $F_{\alpha(k-1,\ n-k)}$ هي القيمة الحرجة لـ F عند مستوى المعنوية α و$(k-1)$ هي درجات حرية البسط و$(n-k)$ درجات حرية المقام. كأسلوب بديل يمكن الحصول على p-value الخاصة بـ F من (13.3.8) وإذا كانت قليلة بشكل كاف، نرفض H_0.

وقبل الانتقال إلى نقطة أخرى، بالعودة إلى مثال 5.7 في الفصل (7). من انحدار (7.10.7) رأينا أن RGDP(GDP النسبي) وRGDP المربعة تفسر فقط 5.3% من التباين في GDPG (معدل نمو الـ GDP) في عينة هذه القيمة معنوية تختلف عن الصفر؟ كيف يمكننا معرفة ذلك؟

تذكر ما سبق واستعرضناه من «العلاقة المهمة بين R^2 وF»، فقد تحدثنا عن العلاقة بين R^2 وF كما في (11.5.8) أو (12.5.8) في حالة خاصة عندما يوجد متغيران اثنان فقط. كما لاحظنا، إذا كانت R^2 مساوية للصفر، فإن F تساوي الصفر أيضاً، وهذه هي الحالة المناظرة لكون المتغيرات المنحدرة ليس لها أي تأثير على المتغير المنحدر عليه. وبالتالي إذا عوضنا عن $R^2 = 0.053$ في المعادلة (12.5.8) نحصل على

$$F = \frac{0.053/2}{(1-0.053)/116} = 3.2475 \qquad \textbf{(13.5.8)}$$

تحت صحة الفرض العدمي القائل بأن $R^2 = 0$، فإن قيمة F تتبع توزيع F بدرجات حرية 2 و116. (لاحظ أنه يوجد 119 مفردة ومتغيران مفسران) من جدول F نرى أن هذه القيمة لـ F معنوية عند مستوى معنوية 5%، وقيمة p-value تساوي 0.0425. وبالتالي نرفض الفرض العدمي القائل بأن هذين المتغيرين ليس لهما أي تأثير على المتغير التابع، بغض النظر عن أن قيمة R^2 تساوي فقط 0.053.

هذا المثال يجعلنا نتطرق إلى ملحوظة عملية مهمة وهي أنه في البيانات المقطعية التي تشتمل على عدد كبير من المشاهدات، عادة ما يتم الحصول على R^2 منخفضة

بسبب الاختلاف أو التنوع الموجود في وحدات البيانات المقطعية . وبالتالي يجب على الباحث ألا ينزعج أو يقلق من الوصول إلى قيمة لـ R^2 منخفضة في انحدار خاص ببيانات مقطعية . النقطة المهمة هي صحة توصيف النموذج ، وأن معاملات الانحدار المقدرة لها الإشارة الصحيحة (بمعنى الإشارة المتوقعة وفقاً للإطار النظري للموضوع) والمرغوب فيه دائماً هو المعنوية الإحصائية لهذه المعاملات المقدرة . وعلى الباحث أن يختبر المعنوية الإحصائية الفردية لكل مقدر في (7.10.7) على حدة عند مستوى معنوية 5% أو عند مستوى معنوية أفضل (بمعنى أقل من 5%) .

المساهمة «الحدية» أو «الزائدة» للمتغير المفسر:

The Incremental or Marginal Contribution of an Explanatory variable

في الفصل (7)، سبق وأن ذكرنا أنه بوجه عام، لا تستطيع توزيع قيمة R^2 بين العديد من المتغيرات المفسرة . نفس مثال وفيات الأطفال، رأينا أن R^2 كانت 0.7077 ولكننا لا نستطيع أن نحدد أي جزء من هذه القيمة يرجع إلى المتغير PGNP وأي منه يرجع إلى معدل حرية المرأة (FLR) حيث هناك إمكانية لوجود ارتباط بين هذين المتغيرين في العينة المتاحة . يمكننا إلقاء مزيد من الضوء على تلك النقطة باستخدام أسلوب تحليل التغاير .

في **مثالنا التوضيحي**، رأينا أن المتغير X_2 (PGNP) منفرداً والمتغير X_3 (FLR) منفرداً لهما معنوية إحصائية على حدة على أساس قيم اختبار t لكل منهما . ورأينا أيضاً أنه على أساس اختبار F التجميعي لكل من المتغيرين ، فإن لدينا معنوية إحصائية لتأثيرها معاً على المتغير Y (وفيات الأطفال) .

الآن دعنا ندخل PGNP وFLR بشكل متتابع للنموذج، بمعنى دعنا أولاً نقوم بعمل انحدار لوفيات الأطفال على PGNP ، ونحدد معنويته ثم نضيف FLR للنموذج لنرى ما إذا كان سيضيف شيئاً ما أم لا (بالطبع ، ترتيب إضافة أي من المتغيرين إلى الآخر يمكن عمله بالعكس) . الإسهام نقصد به هنا الزيادة أو الهامش للمتغير المفسر .

موضوع الإسهام (الزيادة للمتغير المفسر) من المواضيع المهمة عملياً . ففي معظم الأبحاث العملية ، يكون الباحث غير متأكد تماماً من مدى أهمية إضافة متغير X للنموذج، مع العلم بوجود العديد من المتغيرات المفسرة X الموجودة فعلاً في

النموذج. فالباحث لا يرغب في إضافة متغيرات جديدة تساهم بشكل قليل جداً للـ ESS. وبنفس الفكرة، لا يرغب الباحث في استبعاد متغيرات يمكنها بالفعل زيادة ESS. ولكن كيف يمكن للفرد أن يحدد ما إذا كان المتغير X بالفعل سيقلل الـ RSS؟ يمكن بسهولة استخدام تحليل التباين للإجابة على هذا السؤال.

افترض أننا أولاً سنقوم بعمل انحدار لوفيات الأطفال على PGNP ونحصل على الانحدار التالي:

$$\widehat{\text{CM}}_i = 157.4244 - 0.0114\ \text{PGNP} \qquad (14.5.8)$$
$$t = (15.9894) \quad (-3.5156) \qquad r^2 = 0.1662$$
$$p\ \text{value} = (0.0000) \quad (0.0008) \qquad \text{adj}\ r^2 = 0.1528$$

من هذه النتائج، نلاحظ أن PGNP له معنوية إحصائية في التأثير على CM. جدول الـ ANOUA الخاص بهذا الانحدار معطى في الجدول (5.8) بافتراض أن الخطأ u_i يتبع التوزيع الطبيعي، وتحت صحة الفرض القائل بأن PGNP ليس له تأثير على CM، نحصل على قيمة F كالتالي

$$F = \frac{60{,}449.5}{4890.7822} = 12.3598 \qquad (15.5.8)$$

جدول (5.8) جدول ANOVA للانحدار (14.5.8)

Source of variation	SS	df	MSS
ESS (due to PGNP)	60,449.5	1	604,495.1
RSS	303,228.5	62	4890.7822
Total	363,678	63	

هذه القيمة تتبع توزيع F بدرجات حرية 1، 62. هذه القيمة لـ F لها معنوية عالية، حيث إن قيمة p-value تساوي 0.0008. وبالتالي، كما سبق نرفض الفرض العدمي الخاص بأن PGNP ليس له تأثير على CM. لاحظ أن $t^2 = (-3.5156)^2 = 12.32599$ وهي تقريباً نفس قيمة F في (15.5.8). حيث إن قيمة t حصلنا عليها من (14.5.8). ولكن هذا لا يجب أن يكون شيئًا مفاجئًا حيث إن تربيع إحصاء t بدرجات حرية n يساوي قيمة F بدرجات حرية 1 في البسط وn في المقام. وقد تعرضنا لهذه العلاقة من قبل في الفصل (5). لاحظ أنه في المثال الحالي $n = 64$.

وعندما نقوم بانحدار (14.5.8)، دعنا نفترض أننا قررنا إضافة FLR، وبالتالي سنقوم بإجراء انحدار متعدد (1.2.8). **الأسئلة التي نحتاج للإجابة عليها هي :**

1 – ما هي المساهمة الهامشية للـ FLR، مع معلومية أن PGNP موجود بالفعل في النموذج، وهو متغير معنوي بالنسبة لتأثيره على CM؟

2 – هل المساهمة الهامشية لـ FLR لها معنوية إحصائية؟

3 – ما هو معيار إضافة متغيرات جديدة لأي نموذج؟

الأسئلة السابقة يمكن الإجابة عليها من خلال أسلوب الـ ANOVA. للقيام بذلك دعنا نكون جدول (6.8). في هذا الجدول X_2 تشير إلى PGNP و X_3 تشير إلى FLR.

لتحديد المساهمة الزائدية لـ X_3 مع معرفة مساهمة X_2، تكون كالتالي :

$$F = \frac{Q_2/\text{df}}{Q_4/\text{df}}$$
$$= \frac{\text{ESS}_{\text{new}} - \text{ESS}_{\text{old}}/\text{number of new regressors}}{\text{RSS}_{\text{new}}/\text{df}\,(= n - \text{number of parameters in the new model})}$$
$$= \frac{Q_2/1}{Q_4/12} \text{ for our example}$$

(16.5.8)

جدول (6.8) جدول ANOVA لتحديد المساهمة الزائدية للمتغير

Source of variation	SS	df	MSS
ESS due to X_2 alone	$Q_1 = \hat{\beta}_{12}^2 \sum x_2^2$	1	$\frac{Q_1}{1}$
ESS due to the addition of X_3	$Q_2 = Q_3 - Q_1$	1	$\frac{Q_2}{1}$
ESS due to both X_2, X_3	$Q_3 = \hat{\beta}_2 \sum y_i x_{2i} + \hat{\beta}_3 \sum y_i x_{3i}$	2	$\frac{Q_3}{2}$
RSS	$Q_4 = Q_5 - Q_3$	$n-3$	$\frac{Q_4}{n-3}$
Total	$Q_5 = \sum y_i^2$	$n-1$	

جدول (7.8) جدول ANOVA للمثال التوضيحي : التحليل الزائدي

Source of variation	SS	df	MSS
ESS due to PGNP	60,449.5	1	60,449.5
ESS due to the addition of FLR	196,912.9	1	196,912.9
ESS due to PGNP and FLR	257,362.4	2	128,681.2
RSS	106,315.6	61	1742.8786
Total	363,678	63	

بحيث إن ESS = ESS_{new} في إطار النموذج الجديد (أي بعد إضافة المتغير الجديد = Q_3) وتكون ESS = ESS_{old} في إطار النمـوذج القـديـم (=Q_1) وتكون RSS = RSS_{new} في إطار النموذج الجديد (أي بعد الأخذ في الاعتبار جميع المتغيرات المفسرة = Q_4). النتائج الخاصة بالمثال التوضيحي معطاة في جدول (7.8).

والآن عند تطبيق (16.5.8) نحصل على التالي :

$$F = \frac{196{,}912.9}{1742.8786} = 112.9814 \quad \textbf{(17.5.8)}$$

وتحت صحة الفروض التقليدية، فإن هذه القيمة F تتبع توزيع F بدرجات حرية معنوية عالية، مما يدل على أن إضافة FLR للنموذج إضافة بشكل معنوي قيمة الـ ESS وبالتالي قيمة R^2. وبالتالي فإن FLR لابد من إضافته للنموذج.

ومرة أخرى ممكن ملاحظة أنك إذا أخذت قيمة معامل FLR وقمت بتربيعها في نموذج الانحدار المتعدد (1.2.8)، وهي $(-10.6293)^2$ ستحصل على قيمة F الموجودة في (17.5.8) مع تجاهل خطأ التقريب.

وعلى سبيل المصادفة، يمكن أن نلاحظ أن النسبة F الموجودة في (16.5.8) ممكن الحصول عليها باستخدام قيم R^2 فقط، كما فعلنا من قبل في (13.5.8). وكما سنرى في تمرين 2.8، فإن النسبة F المعطاة في (16.5.8) مساوية للنسبة F التالية[9] :

$$\begin{aligned} F &= \frac{(R^2_{\text{new}} - R^2_{\text{old}})/\text{df}}{(1 - R^2_{\text{new}})/\text{df}} \\ &= \frac{(R^2_{\text{new}} - R^2_{\text{old}})/\text{number of new regressors}}{(1 - R^2_{\text{new}})/\text{df}\,(= n - \text{number of parameters in the new model})} \end{aligned} \quad \textbf{(18.5.8)}$$

هذه النسبة F تتبع توزيع F بدرجات حرية البسط والمقام المناسبة وهي 1 و 61 بالترتيب في مثالنا التوضيحي.

في مثالنا الحالي، $R^2_{\text{new}} = 0.7077$ [من المعادلة (1.2.8)] و $R^2_{\text{old}} = 0.1662$ [من المعادلة (14.5.8)]. وبالتالي

(9) هذه القيمة للـ F تعتبر حالة خاصة من الحالات الأكثر عمومية لاختبار F والمعطاة في (9.7.8) أو (10.7.8) في الفقرة 7.8 .

$$F = \frac{(0.7077 - 0.1662)/1}{(1 - 0.7077)/61} = 113.05 \qquad (19.5.8)$$

وهذه القيمة مساوية تقريباً للقيمة التي حصلنا عليها من قبل من (17.5.8) مع استثناء خطأ التقريب. هذه القيمة للـ F تعتبر قيمة عالية المعنوية، مؤكدة مرة أخرى على أهمية إضافة المتغير FLR للنموذج المقترح.

ملاحظة مهمة: إذا كنت ستستخدم R^2 للحصول على قيمة اختبار F كما هو معطى في (11.5.8) تأكد أن المتغير التابع واحد في النموذج القديم والنموذج الجديد. إذا كانوا مختلفين فاستخدام صيغة (16.5.8) للحصول على قيمة اختبار F.

متى يمكن إضافة متغير واحد جديد؟ طريقة اختبار F السابق ذكرها تعتبر طريقة نظرية للتحديد، إضافة متغير جديد لنموذج الانحدار من عدمه. فدائماً إمكانية ما يواجهه الباحثون مشكلة الاختبار من بين عدد من النماذج المتنافسة والخاصة بنفس المتغير التابع، ولكن كلاً منهما يحتوي على متغيرات مفسرة مختلفة. وكنوع من أنواع الاختبار (خصوصاً وكثيراً ما تكون الخلفية النظرية للتحليل ضعيفة) يلجأ الباحثون إلى اختبار النموذج الذي له أعلى قيمة R^2 المعدلة. وبالتالي إذا كانت إضافة متغير جديد تزيد من قيمة$\overline{R}^2$، فيتم الاحتفاظ به في النموذج، رغم أنه قد لا يقلل RSS بشكل معنوي بالمعنى الإحصائي المتعارف عليه. وبالتالي يكون السؤال كالتالي: متى تزيد قيمة R^2 المعدلة؟ يمكن إثبات أن قيمة $\overline{R}^2$ ستزيد عندما تزداد القيمة المطلقة t الخاصة بمعامل المتغير الذي تمت إضافته عن الواحد الصحيح، وذلك عندما تكون فيه t محسوبة تحت صحة الفرض العدمي الخاص بعدم معنوية معامل المتغير في المجتمع. [وهذه هي قيمة t الموجودة في (2.3.5) تحت صحة الفرض العدمي والخاص بأن قيمة β الحقيقية تساوي الصفر][10].

ومن الممكن صياغة الطريقة السابقة بشكل مختلف كالتالي: ستزداد قيمة$\overline{R}^2$مع كل إضافة لمتغير مفسر جديد إذا كانت قيمة $F(=t^2)$ الخاصة لهذا المتغير تزيد عن 1.

وبتطبيق ذلك، نجد أن المتغير FLR في مثالنا الخاص بوفيات الأطفال والذي له قيمة t تساوي 10.6293– أو قيمة F تساوي 112.9814 يزيد من قيمة$\overline{R}^2$. وهذه القيمة بالفعل تزداد عند إضافة FLR للنموذج، فهي تزيد من 0.1528 إلى 0.6981.

(10) للإثبات النظري، انظر في Dennis J. Aigner, Basic Econometrics, Prentice Hall, Englewood Cliffs, N. J., 1971, pp. 91–92.

متى يمكن إضافة مجموعة من المتغيرات؟ هل ممكن أن نحصل على قاعدة مماثلة يمكن استخدامها لتحديد إمكانية إضافة (أو حذف) مجموعة من المتغيرات في نموذج ما من عدمه؟ الإجابة يجب أن تكون واضحة من (18.5.8) : إذا كانت إضافة (أو حذف) مجموعة من المتغيرات من نموذج ما تعطي قيمة أعلى لـ F (أقل) من 1، فإن R^2 ستزداد (ستقل).

وبالطبع، من (18.5.8) يمكن للقارئ بسهولة أن يحدد ما إذا كانت الإضافة (أو الحذف) لمجموعة من المتغيرات ستزيد بشكل معنوي (أو ستقل) قوة المتغيرات المفسرة في نموذج الانحدار.

6.8 اختبار تساوي معاملي انحدار :
TESTING THE EQUALITY OF TWO REGRESSION COEFFICIENTS

اعتبر نموذج الانحدار المتعدد التالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i \qquad \textbf{(1.6.8)}$$

إذا أردنا اختبار التالي :

$$H_0: \beta_3 = \beta_4 \quad \text{or} \quad (\beta_3 - \beta_4) = 0 \qquad \textbf{(2.6.8)}$$
$$H_1: \beta_3 \neq \beta_4 \quad \text{or} \quad (\beta_3 - \beta_4) \neq 0$$

أي نريد اختبار تساوي معاملات الميل β_3 و β_4 معاً .

الفرض العدمي السابق هو فرض مهم من الناحية العملية . فعلى سبيل المثال ، اعتبر أن (1.6.8) تمثل دالة الطلب على سلعة ما، حيث Y = الكمية المطلوبة من السلعة ، X_2 = سعر السلعة ، X_3 = دخل المستهلك و X_4 = ثروة المستهلك . وبالتالي يكون الفرض العدمي في هذه الحالة، يعني تساوي معامل الدخل مع معامل الثروة . أو إذا كانت Y_i و X's معبر عنهما في الصورة اللوغاريتمية ، فإن الفرض العدمي في (2.6.8) سيعني أن مرونة الدخل ومرونة الثروة للاستهلاك متساويتان . (لماذا؟)

كيف يمكن أن نختبر مثل هذا الفرض؟ تحت صحة الفروض التقليدية يمكن إثبات أن :

$$t = \frac{(\hat{\beta}_3 - \hat{\beta}_4) - (\beta_3 - \beta_4)}{\text{se}\,(\hat{\beta}_3 - \hat{\beta}_4)} \qquad \textbf{(3.6.8)}$$

هذه القيمة تتبع توزيع t بدرجات حرية $(n-4)$، حيث إن النموذج (1.6.8) هو نموذج يحتوي على أربعة متغيرات، وبالتالي، فإنه عموماً تكون درجات حرية t هي $(n-k)$ و k تمثل إجمالي عدد المعالم المقدرة، مضاف إليها الجزء الثابت. قيمة $se(\hat{\beta}_3 - \hat{\beta}_4)$ يمكن الحصول عليها من المعادلة المعروفة (انظر ملحق **A** للتفصيل).

$$se(\hat{\beta}_3 - \hat{\beta}_4) = \sqrt{var(\hat{\beta}_3) + var(\hat{\beta}_4) - 2\,cov(\hat{\beta}_3, \hat{\beta}_4)} \quad (4.6.8)$$

إذا استبدلنا المعادلة السابقة لـ $se(\beta_3 - \beta_4)$ في المعادلة (3.6.8) ستكون قيمة إحصاء الاختبار كالتالي :

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{var(\hat{\beta}_3) + var(\hat{\beta}_4) - 2\,cov(\hat{\beta}_3, \hat{\beta}_4)}} \quad (5.6.8)$$

والآن يمكن تلخيص خطوات الاختبار في الخطوات التالية :

1 - قدر β_3 و β_4. يمكن لأي حزمة حاسب آلي إحصائية القيام بذلك.

2 - معظم هذه الحزم الإحصائية تحسب بشكل أوتماتيكي التباين والتغاير بين المعالم المقدرة[11].

3 - احصل على النسبة t من (5.6.8). لاحظ أن الفرض العدمي في الوضع الحالي هو $(\beta_3 - \beta_4) = 0$.

4 - إذا كانت قيمة t المحسوبة في (5.6.8) تزيد عن القيمة الحرجة للـ t عند نفس مستوى المعنوية ودرجات الحرية المحددة، فإنه يمكن رفض الفرض العدمي. وبخلاف ذلك، لايمكن رفض هذا الفرض. وكطريقة بديلة إذا كانت القيمة p (*p*-value) لإحصاء الاختبار t الموجود في (5.6.8) منخفضة بشكل كاف، يمكن رفض الفرض العدمي. لاحظ أنه كلما قلت *p*-value كلما كان ذلك دليلاً أكبر ضد الفرض العدمي.

وبالتالي، عندما نقول إن قيمة *p*-value صغيرة أو صغيرة بشكل مقبول، نعني أنها أقل من مستوى المعنوية وأي 10، 5 أو 1%.

لاحظ أنه أحياناً يكون للأحكام الشخصية دور في اتخاذ القرار.

(11) الصيغة الجبرية للتغاير يتم استخدامها في هذا الإطار. ملحق 2 يعطي صياغة مختصرة لهذا التغاير باستخدام رموز المصفوفات.

مثال 2.8

مرة أخرى - دالة التكلفة التكعيبية : **The Cubic Cost Function Revisited**

بالعودة إلى دالة التكلفة التكعيبية المقدرة في الفقرة 10.7، والموجودة في المعادلة التالية :

$$\hat{Y}_i = 141.7667 + 63.4777X_i - 12.9615X_i^2 + 0.9396X_i^3$$
$$\text{se} = (6.3753) \quad (4.7786) \quad (0.9857) \quad (0.0591) \qquad (6.10.7)$$
$$\text{cov}(\hat{\beta}_3, \hat{\beta}_4) = -0.0576; \qquad R^2 = 0.9983$$

حيث : Y هي التكلفة الكلية ، و X هي الإنتاج والأرقام داخل الأقواس تمثل الأخطاء المعيارية المقدرة .

افترض أننا نريد اختبار أن معامل X^2 ومعامل X^3 في معادلة التكلفة التكعيبية متساويان . أي أن $(\beta_3 = \beta_4)$ أو $(\beta_3 - \beta_4 = 0)$. في معادلة الانحدار (6.10.7) لدينا كل المدخلات المطلوبة للقيام باختبار t الموجودة في (5.6.8) . وذلك من خلال الخطوات التالية :

$$t = \frac{\hat{\beta}_3 - \hat{\beta}_4}{\sqrt{\text{var}(\hat{\beta}_3) + \text{var}(\hat{\beta}_4) - 2\,\text{cov}(\hat{\beta}_3, \hat{\beta}_4)}}$$
$$= \frac{-12.9615 - 0.9396}{\sqrt{(0.9867)^2 + (0.0591)^2 - 2(-0.0576)}} \qquad (6.6.8)$$
$$= \frac{-13.9011}{1.0442} = -13.3130$$

يمكن للقارئ أن يثبت أن درجات الحرية = 6 (لماذا؟) قيمة t المحسوبة تزيد عن القيمة الحرجة للـ t حتى عند مستوى معنوية 0.002 (0.02%) (اختبار ذو طرفين)، قيمة (p-value) صغيرة للغاية 0.000006 وبالتالي نرفض الفرض العدمي الخاص بتساوي معاملات X^2 و X^3 في دالة التكلفة التكعيبية .

7.8 المربعات الصغرى المقيدة.. اختبار قيود التساوي الخطي: RESTRICTED LEAST SQUARES' TESTING LINEAR EQUALITY RESTRICTIONS:

في بعض الأحيان تقترح النظرية الاقتصادية بعض قيود التساوي الخطية لمعاملات نموذج الانحدار. فعلى سبيل المثال، دعنا نعتبر دالة Cobb - Douglas للإنتاج التالية :

$$Y_i = \beta_1 X_{2i}^{\beta_2} X_{3i}^{\beta_3} e^{u_i} \qquad (1.9.7) = (1.7.8)$$

حيث : Y = الإنتاج، X_2 = مدخل العمالة، X_3 = مدخل رأس المال إذا تمت إعادة كتابة الدالة السابقة في صورة لوغاريتمية تكون كالتالي :

$$\ln Y_i = \beta_0 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \qquad (2.7.8)$$

حيث : $\beta_0 = \ln \beta_1$

والآن مع افتراض ثبات العائد إلى القياس (التغير النسبي في الناتج نسبة للتغير النسبي في المدخلات)، النظرية الاقتصادية تقترح التالي :

$$\beta_2 + \beta_3 = 1 \qquad (3.7.8)$$

وهذا يعتبر مثالاً لقيد تساوي خطي[12].

كيف يمكن التحقق من فرض ثبات العائد إلى القياس أي تحقق القيد (3.7.8)؟ يمكن القيام بذلك من خلال الأسلوبين التاليين :

أسلوب اختبار t : The t - test Approach

الطريقة الأسهل هي تقدير (2.7.8) بالطريقة التقليدية بدون وضع في الاعتبار القيد (3.7.8). وهذا يسمى الانحدار غير المقيد. ويعد تقدير β_2، β_3 (مثلاً باستخدام طريقة المربعات الصغرى العادية). نقوم بعمل اختبار للفرض أو القيد (3.7.8) باستخدام t-test (3.6.8) كالتالي :

$$t = \frac{(\hat{\beta}_2 + \hat{\beta}_3) - (\beta_2 + \beta_3)}{\text{se}\,(\hat{\beta}_2 + \hat{\beta}_3)}$$

$$= \frac{(\hat{\beta}_2 + \hat{\beta}_3) - 1}{\sqrt{\text{var}\,(\hat{\beta}_2) + \text{var}\,(\hat{\beta}_3) + 2\,\text{cov}\,(\hat{\beta}_2 \hat{\beta}_3)}} \qquad (4.7.8)$$

حيث إنه **تحت صحة الفرض العدمي**، فإن $\beta_2 + \beta_3 = 1$، وبالتالي يكون المقام هو الخطأ المعياري لـ $\beta_3 + \beta_3$. وباتباع الفقرة (6.8)، إذا كانت قيمة t المحسوبة في (4.7.8) تزيد عن قيمة t الحرجة عن نفس مستوى المعنوية، فإننا نرفض الفرض الخاص بثبات العائد إلى القياس، وبخلاف ذلك لا نرفض الفرض العدمي.

أسلوب اختبار F.. المربعات الصغرى المقيدة :

The F-test Approach: Restricted least squares

طريقة اختبار t السابقة تعتبر اختباراً تالياً لعملية التقدير، فنحن نحاول معرفة ما إذا كان القيد الخطي متحققاً أم لا بعد تقدير نموذج الانحدار غير المقيد. الأسلوب

(12) إذا كان لدينا $\beta_2 + \beta_3$. فإن هذه العلاقة تعتبر مثالاً لقيد عدم تساوي خطي للتعامل مع مثل هذا القيد ، نحتاج استخدام أحد أساليب البرمجة الرياضية .

المباشر للقيام بهذا الاختبار يكون من خلال إدراج القيد (3.7.8) في عملية التقدير نفسها. في المقال الحالي، هذه الطريقة يمكن عملها بسهولة كالتالي. من (3.7.8) نلاحظ أن :

$$\beta_2 = 1 - \beta_3 \quad \textbf{(5.7.8)}$$

أو

$$\beta_3 = 1 - \beta_2 \quad \textbf{(6.7.8)}$$

وبالتالي استخدام أي من المعادلتين السابقتين يجعلنا نتجاهل أحد المعاملات β من (2.7.8) وتقدير المعادلة الناتجة. وبالتالي إذا اسخدمنا (5.7.8) يمكن كتابة دالة إنتاج Cobb - Douglas كالتالي :

$$\ln Y_i = \beta_0 + (1 - \beta_3)\ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

$$= \beta_0 + \ln X_{2i} + \beta_3 (\ln X_{3i} - \ln X_{2i}) + u_i$$

أو

$$(\ln Y_i - \ln X_{2i}) = \beta_0 + \beta_3(\ln X_{3i} - \ln X_{2i}) + u_i \quad \textbf{(7.7.8)}$$

أو

$$\ln\left(\frac{Y_i}{X_{2i}}\right) = \beta_0 + \beta_3 \ln\left(\frac{X_{3i}}{X_{2i}}\right) + u_i \quad \textbf{(8.7.8)}$$

حيث : $\left(\frac{Y_i}{X_{2i}}\right)$ = نسبة الناتج إلى العمالة و $\left(\frac{X_{3i}}{X_{2i}}\right)$ = نسبة رأس المال إلى العمالة. وهاتان النسبتان لهما أهمية كبيرة من الناحية الاقتصادية.

لاحظ كيف تم تحويل المعادلة (2.7.8) فبمجرد تقدير β_3 من (7.7.8) أو (8.7.8)، β_2 يمكن تقديرها بسهولة من العلاقة (5.7.8). وهذه الطريقة تضمن أن يكون مجموع المعاملات المقدرة لهذين المدخلين يساوي الواحد الصحيح. هذه الطريقة الموضحة في (7.7.8) و(8.7.8) معروفة باسم طريقة المربعات الصغرى المقيدة Restricted least Squanes (RLS). وهذه الطريقة يمكن تعميمها واستخدامها مع نماذج محتوى على أي عدد من المتغيرات المفسرة وأكثر من قيد خطي واحد. التعميم الخاص بهذه الطريقة موضح بالتفصيل في Theil.[13] (انظر اختبار F العام التالي)

(1) Henri Theil, Principles of Econometrics, John Wiley & Sons, New York, 1971, pp. 43–45.

السؤال الآن هو : كيف تم حساب مقدرات المربعات الصغرى للانحدار المقيدة وغير المقيدة؟ أو كصياغة أخرى للسؤال : كيف نتأكد من صحة المقدرات المقيدة؟ للإجابة على هذا السؤال، دعنا نطبق اختبار F كالتالي : دع

$\sum \hat{u}^2_{\text{UR}}$ = تمثل RSS للانحدار غير المقيد (2.7.8)

$\sum \hat{u}^2_{\text{R}}$ = تمثل RSS للانحدار المقيد (7.7.8)

m = عدد القيود الخطية (قيد واحد في المثال الحالي)

k = عدد المعالم في الانحدار غير المقيد.

n = عدد المشاهدات

وبالتالي :

$$F = \frac{(\text{RSS}_{\text{R}} - \text{RSS}_{\text{UR}})/m}{\text{RSS}_{\text{UR}}/(n-k)} = \frac{\left(\sum \hat{u}^2_{\text{R}} - \sum \hat{u}^2_{\text{UR}}\right)/m}{\sum \hat{u}^2_{\text{UR}}/(n-k)} \tag{9.7.8}$$

هذه القيمة تتبع توزيع F بدرجات حرية m، $(n-k)$. (لاحظ أن : UR ترمز إلى الطريقة غير المقيدة والـ R ترمز إلى الطريقة المقيدة).

اختبار F السابق ذكره يمكن التعبير عنه في صورة R^2 كالتالي :

$$F = \frac{(R^2_{\text{UR}} - R^2_{\text{R}})/m}{(1 - R^2_{\text{UR}})/(n-k)} \tag{10.7.8}$$

حيث : R^2_{UR} و R^2_{R} هما قيم R^2 التي تم الحصول عليها بالانحدار غير المقيد، والانحدار المقيد على الترتيب، أي أنه من الانحدار (2.7.8) و (7.7.8) يجب أن نلاحظ أن :

$$R^2_{\text{UR}} \geq R^2_{\text{R}} \tag{11.7.8}$$

و

$$\sum \hat{u}^2_{\text{UR}} \leq \sum \hat{u}^2_{\text{R}} \tag{12.7.8}$$

في التمرين 4.8 هناك سؤال للقارئ للتأكد من صحة العلاقات السابقة.

ملحوظة مهمة : عند استخدام (10.7.8) يجب أن يضع القارئ في الاعتبار أنه في

حالة أن يكون المتغير التابع الموجود في النموذج المقيد مختلفًا عن نظرية في النموذج غير المقيد سيكون من الخطأ مقارنة R^2_{UR} و R^2_R مباشرة.

في مثل هذه الحالة، استخدم الأسلوب الذي استعرضناه من قبل في الفصل 7 لمقارنة هاتين القيمتين (انظر مثال 3.8 التالي) أو كطريقة أخرى للتعامل مع هذه الحالة، يمكن استخدام اختبار F المعطى في (9.7.8).

مثال 3.8

دالة الإنتاج لـ Cobb-Douglas للاقتصاد المكسيكي في الفترة 1955 – 1974

The Cobb - Dauglas Production Function for the Mexican Ecanomy, 1955-1974

لتوضيح الأسلوب السابق مناقشته، دعنا نعتبر البيانات المعطاة في جدول (8.8). وكمحاولة لتوفيق دالة إنتاج Cobb - Douglas لهذه البيانات حصلنا على النتائج التالية:

$$\widehat{\ln GDP}_t = -1.6524 + 0.3397 \ln Labor_t + 0.8460 \ln Capital_t \qquad (13.7.8)$$

$$t = (-2.7259) \quad (1.8295) \quad (9.0625)$$

$$p \text{ value} = (0.0144) \quad (0.0849) \quad (0.0000)$$

$$R^2 = 0.9951 \qquad RSS_{UR} = 0.0136$$

حيث : RSS_{UR} هي RSS غير المقيدة، حيث إننا لا نفترض أى قيود عند القيام بعملية التقدير كما في (13.7.8).

سبق وأن رأينا في الفصل (7)، كيف يمكن تفسير معاملات دالة إنتاج Cobb-Douglas. كما رأينا فقد كانت مرونة الناتج / العمالة حوالي 0.34 ومرونة الناتج / رأس المال حوالي 0.85. إذا جمعنا هذين المعاملين سنحصل على 1.19 مما يعني أنه من المحتمل أن تكون هناك زيادة في العوائد خلال هذه الفترة في الاقتصاد المكسيكي. وبالطبع لا نستطيع أن نحدد ما إذا كان هناك فرق بين هذه القيمة والواحد الصحيح وهو فرق معنوي.

وللتحقق من ذلك، دعنا نلتزم بقيد ثبات العائد والذي يعطينا الانحدار التالي:

$$\widehat{\ln (GDP/Labor)}_t = -0.4947 + 1.0153 \ln (Capital/Labor)_t \qquad (14.7.8)$$

$$t = (-4.0612) \quad (28.1056)$$

$$p \text{ value} = (0.0007) \quad (0.0000)$$

$$R^2_R = 0.9777 \qquad RSS_R = 0.0166$$

حيث : RSS_R هو الـ RSS المقيد حيث وضعنا القيد الخاص بثبات العوائد.

وبما أن المتغير التابع ليس متغيراً واحداً في الانحدارين السابقين، يجب أن نستخدم اختبار F المعطى في (9.7.8) لدينا البيانات المطلوبة للحصول على قيمة F كالتالي:

$$F = \frac{(RSS_R - RSS_{UR})/m}{RSS_{UR}/(n-k)}$$

$$= \frac{(0.0166 - 0.0136)/1}{(0.0136)/(20-3)}$$

$$= 3.75$$

لاحظ أنه في مثالنا الحالي ، فإن $m = 1$، حيث إننا وضعنا قيداً واحداً فقط و$(n - k)$ تساوي 17 ، حيث إن لدينا 20 مفردة وثلاث معلمات في الانحدار غير المقيد .

قيمة F تتبع توزيع F بدرجة حرية 1 في البسط و17 في المقام . ويمكن للقارئ بسهولة التحقق من أن هذه القيمة للـ F غير معنوية عند مستوى المعنوية 5% (انظر ملحق D، جدول D3) .

وبالتالي ، فيمكن أن نستنتج الآن أن الاقتصاد المكسيكي من المحتمل أنه يتميز بثبات العائد خلال هذه الفترة الزمنية ، وبالتالي لا ضرر من استخدام الانحدار المقيد المعطى في (14.7.8) ومن نتائج الانحدار يمكن ملاحظة أنه إذا زادت نسبة رأس المال/ العملة بـ 1% ففي المتوسط تزداد إنتاجية العمالة بحوالي 1% .

جدول (8.8) GDP الحقيقي ، التوظيف ورأس المال الحقيقي الثابت – المكسيك

Real GDP, Emplorymen, and Real fixed Capital - Mexico

Year	GDP*	Employment†	Fixed capital‡
1955	114043	8310	182113
1956	120410	8529	193749
1957	129187	8738	205192
1958	134705	8952	215130
1959	139960	9171	225021
1960	150511	9569	237026
1961	157897	9527	248897
1962	165286	9662	260661
1963	178491	10334	275466
1964	199457	10981	295378
1965	212323	11746	315715
1966	226977	11521	337642
1967	241194	11540	363599
1968	260881	12066	391847
1969	277498	12297	422382
1970	296530	12955	455049
1971	306712	13338	484677
1972	329030	13738	520553
1973	354057	15924	561531
1974	374977	14154	609825

*Millions of 1960 pesos;
†Thousands of people;
‡Millions of 1960 pesos.
Source: Victor J. Elias, *Sources of Growth: A Study of Seven Latin American Economies,* International Center for Economic Growth, ICS Press, San Francisco, 1992. Data from Tables E5, E12, and E14.

اختبار F العام[14] : Reneral F testing

اختبار F المعطى في (10.7.8) أو المكافئ له في (9.7.8) يعطي طريقة عامة لاختبار معنوية واحد أو أكثر من المعالم في نموذج انحدار تحتوي على k متغير كالتالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad \textbf{(15.7.8)}$$

اختبار F المعطى في (16.5.8)، أو اختبار t في (3.6.8) ليس إلا حالة خاصة من (10.7.8). وبالتالي فروض عدمية كالتالي :

$$H_0 : \beta_2 = \beta_3 \qquad \textbf{(16.7.8)}$$

$$H_0 : \beta_3 + \beta_4 + \beta_5 = 3 \qquad \textbf{(17.7.8)}$$

والتي تمثل بعض القيود الخطية على معالم النموذج الذي يحتوي على k متغير. أو الفروض التالية :

$$H_0 : \beta_3 = \beta_4 = \beta_5 = \beta_6 = 3 \qquad \textbf{(18.7.8)}$$

والتي تعني غياب تأثير بعض المتغيرات المفسرة في النموذج، يمكن اختيارها باستخدام اختبار F في (10.7.8).

من المناقشة الموجودة في الفقرة 5.8 و 7.8، يمكن أن يلاحظ القارئ أن الأسلوب العام لاختبار F هو كالتالي: هناك نموذج كبير أو نموذج غير مقيد (15.7.8) ثم هناك نموذج أصغر (النموذج المقيد) والذي يتم الحصول عليه من النموذج الأكبر بعد حذف بعض المتغيرات، كما في (18.7.8)، أو بعد إضافة بعض القيود الخطية على واحد أو أكثر من معاملات النموذج الأكبر، كما في (16.7.8) أو (17.7.8).

ثم نوفق النموذج المقيد وغير المقيد للبيانات، ونحصل على معاملات التحديد الخاصة بكل نموذج وهي R^2_{UR} و R^2_{R}.

ونلاحظ أن درجات الحرية في النموذج غير المقيد (تساوي $n - k$)، أما في النموذج المقيد (تساوي m)، m هي عدد القيود الخطية [على سبيل المثال، قيد واحد

(14) إذا كنا نستخدم طريقة الإمكان الأعظم للتقدير، فهناك اختبار مناظر للاختبار السابق مناقشته هو اختبار الإمكان الأعظم ونظراً لتعقيده بعض الشيء فهو مشروح في ملحق الفصل. ولمناقشة أعمق، انظر Theil، op.cit.، pp. 179–184.

فقط في (16.7.8) أو (18.7.8)] أو عدد المتغيرات المحذوفة من النموذج [مثال $m = 4$ في (18.7.8) حيث أربعة متغيرات تعتبر متغيبة من النموذج). ثم نسحب النسبة F كما هو موضح في (9.7.8) أو (10.7.8) وتستخدم القاعدة التالية لأخذ القرار: إذا زادت F المحسوبة عن $F_\alpha(m, n-k)$ [حيث $F_\alpha(m, n-k)$ هي F الحرجة عند مستوى المعنوية α] نرفض الفرض العدمي، وبخلاف ذلك لا نرفض الفرض العدمي.

دعنا نستعرض المثال التالي للتوضيح:

مثال 4.8

الطلب على الدجاج في الولايات المتحدة خلال الفترة 1960 – 1982
The Demand for chicken in the United states, 1960–1982

في تمرين 19.7، أحد الأسئلة كان يعتبر الدالة التالية للطلب على النجاح:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_4 \ln X_{4t} + \beta_5 \ln X_{5t} + u_i \qquad \textbf{(19.7.8)}$$

حيث Y: = استهلاك الدجاج للفرد مقاس بالرطل، X_2 = الدخل الحقيقي للفرد مقاس بالدولار، X_3 = سعر الدجاج بالرطل بالسنت، X_4 = سعر لحم الخنزير بالرطل مقاس بالسنت و X_5 = سعر اللحوم بالرطل مقاس بالسنت.

في هذا النموذج β_2، β_3، β_4 وβ_5 تمثل المرونات التالية (بالترتيب)، مرونة الدخل، مرونة سعر الدجاج، مرونة سعر لحم الخنزير مقارنة بسعر الدجاج، مرونة سعر اللحوم مقارنة بسعر الدجاج. (لماذا) وفقاً للنظرية الاقتصادية لدينا التالي:

$\beta_2 > 0$

$\beta_3 < 0$

$\beta_4 > 0$، إذا اعتبرنا الدجاج ولحم الخنزير سلعًا تنافسية

< 0، إذا اعتبرنا الدجاج ولحم الخنزير سلعًا مكملة **(20.7.8)**

$= 0$، إذا اعتبرنا الدجاج ولحم الخنزير سلعًا غير مرتبطة

$\beta_5 > 0$، إذا اعتبرنا الدجاج واللحوم سلعًا تنافسية

< 0، إذا اعتبرنا الدجاج واللحوم سلعًا مكملة

$= 0$، إذا اعتبرنا الدجاج واللحوم سلعًا غير مرتبطة

إذا افترض الشخص أن الدجاج ولحم الخنزير واللحوم سلع غير مرتبطة، فإن استهلاك الدجاج لا يتأثر بسعر لحم الخنزير أو اللحوم.

وبالتالي باختصار فإن:

$$H_0: \beta_4 = \beta_5 = 0 \qquad \textbf{(21.7.8)}$$

وبالتالي الانحدار المقيد يصبح كالتالي :

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_i \qquad (22.7.8)$$

معادلة (19.7.8) هي بالطبع خاصة بالانحدار غير المقيد. باستخدام البيانات المعطاة في تمرين 19.7، نحصل على التالي :

الانحدار غير المقيد

$$\widehat{\ln Y_t} = \underset{(0.1557)}{2.1898} + \underset{(0.0833)}{0.3425} \ln X_{2t} - \underset{(0.1109)}{0.5046} \ln X_{3t} + \underset{(0.0997)}{0.1485} \ln X_{4t} + \underset{(0.1007)}{0.0911} \ln X_{5t}$$

$$R^2_{\text{UR}} = 0.9823 \qquad (23.7.8)$$

الانحدار المقيد

$$\widehat{\ln Y_t} = \underset{(0.1162)}{2.0328} + \underset{(0.0247)}{0.4515} \ln X_{2t} - \underset{(0.0635)}{0.3772} \ln X_{3t}$$

$$R^2_{\text{R}} = 0.9801 \qquad (24.7.8)$$

حيث إن الأرقام الموجودة بين الأقواس هي تقديرات الأخطاء القياسية.

لاحظ أن : قيمة R^2 الموجودة في (23.7.8) و(24.7.8) يمكن مقارنتهما معاً، حيث إن المتغير التابع هو نفس المتغير في النموذجين.

والآن قيمة النسبة F اللازمة لإجراء اختبار (21.7.8) هي كالتالي :

$$F = \frac{(R^2_{\text{UR}} - R^2_{\text{R}})/m}{(1 - R^2_{\text{UR}})/(n-k)} \qquad (10.7.8)$$

قيم m في المقال الحالي تساوي 2، حيث إن لدينا قيدين اثنين وهما $\beta_4 = 0$ ، $\beta_5 = 0$. درجات حرية المقام هي $(n-k)$ تساوي 18 حيث $n = 23$ و $k = 5$ (حيث يوجد β_5 من المعاملات).

وبالتالي فإن قيمة F هي :

$$F = \frac{(0.9823 - 0.9801)/2}{(1 - 0.9823)/18}$$

$$= 1.1224 \qquad (25.7.8)$$

هذه القيمة تتبع توزيع F بدرجات حرية تساوي 2 و 18.

عند مستوى معنوية مساوي (5%) هذه القيمة لـ F ليست معنوية إحصائياً $[F_{0.5}(2, 18) = 3.55]$. قيمة p-value هي 0.3472. وبالتالي لا يوجد سبب يجعلنا نرفض الفرض العدمي – وهو الفرض القائل بأن الطلب على الدجاج لا يعتمد على أسعار اللحوم أو لحم الخنزير. وبالتالي في عبارات مختصرة يمكن القول بأننا نقبل الانحدار المقيد (24.7.8) كممثل لدالة الطلب على الدجاج.

لاحظ أن دالة الطلب مستوفية لتوقعات اقتصادية مسبقة، وهي خاصة بأن مرونة سعر الدجاج سالبة، ومرونة الدخل موجبة. عموماً فإن المرونة السعرية المقدرة، كقيمة مطلقة، أقل إحصائياً عن الواحد الصحيح، فبالتالي بالرغم من أنها موجبة، إلا أنها إحصائياً أقل من الواحد الصحيح. مما يعني أن الدواجن ليست سلعة رفاهية، حيث إن السلعة يقال عليها سلعة كمالية أو رفاهية إذا كانت مرونة الدخل الخاصة بها أكبر من الواحد الصحيح.

8.8 اختبار استقرار المعلمات أو الاستقرار الهيكلي لنماذج الانحدار: اختبار CHOW

TESTING FOR STRUCTURAL OR PARAMETER STABILITY OF REGRESION MODELS: THE CHOW TEST

عندما نستخدم نموذج انحدار على بيانات سلسلة زمنية، يمكن أن يحدث تغير هيكلي في العلاقة بين المتغير التابع y والمتغيرات المفسرة. والمقصود من التغير الهيكلي أن قيمة معالم النموذج لا تبقى كما هي خلال كل الفترة الزمنية. أحياناً يحدث التغيير الهيكلي نتيجة قوة خارجية (مثل عقوبات الزيت المفروضة من منظمة الأوبك في 1973 و 1979 أو حرب الخليج في 1990 – 1991) أو نتيجة تغير السياسات (مثل التحول من نظام تحويل العملة الثابت إلى نظام تحويل العملة المرن في 1973) وممكن أن يكون سبب التغير الهيكلي قرارات تم اتخاذها من الكونجرس (مثل قوانين الضرائب المقررة بواسطة الرئيس ريجان في فترتي الرئاسة الخاصة به، أو تغير الحد الأدنى للأجور) أو أي أسباب أخرى.

كيف يمكن لنا أن نعرف إذا ما حدث تغير هيكلي؟

للإجابة بالتحديد على هذا السؤال، دعنا نستعرض البيانات المعطاة في جدول (9.8). هذا الجدول يعطي بيانات عن الدخل الشخصي، والمدخرات الشخصية، مقدرة بالبليون دولار، للولايات المتحدة الأمريكية في الفترة 1970 – 1995. دعنا نفترض أننا نريد تقدير دالة الادخار البسيطة والتي تربط بين الادخار (Y)، والدخل الشخصي DPI(X). بما أن البيانات متاحة لدينا، يمكن أن نحصل على مقدرات المربعات الصغرى للانحدار الخاص بـY على X. ولكن إذا قمنا بذلك، نحن نفترض أن العلاقة بين الادخار وDPI لن تتغير خلال فترة الـ 26 عامًا. وهذا قد يكون افتراضًا يصعب تصديقه. فعلى سبيل المثال، من المعروف أنه في 1982 عاشت الولايات المتحدة الأمريكية أسوأ فترات السلام.

جدول (9.8) الدخل الشخصي والادخار الشخصي (بمليون دولار) الولايات المتحدة 1970 – 1995

Savings and personal disposable income (Billions of Dollars), United

Observation	Savings	Income	Observation	Savings	Income
1970	61.0	727.1	1983	167.0	2522.4
1971	68.6	790.2	1984	235.7	2810.0
1972	63.6	855.3	1985	206.2	3002.0
1973	89.6	965.0	1986	196.5	3187.6
1974	97.6	1054.2	1987	168.4	3363.1
1975	104.4	1159.2	1988	189.1	3640.8
1976	96.4	1273.0	1989	187.8	3894.5
1977	92.5	1401.4	1990	208.7	4166.8
1978	112.6	1580.1	1991	246.4	4343.7
1979	130.1	1769.5	1992	272.6	4613.7
1980	161.8	1973.3	1993	214.4	4790.2
1981	199.1	2200.2	1994	189.4	5021.7
1982	205.5	2347.3	1995	249.3	5320.8

حيث وصلت معدلات البطالة إلى 9.6%، وهي الأكبر من 1989، حدث مثل ذلك قد يغير العلاقة بين الادخار وDPI. لدراسة ذلك، دعنا نقسم البيانات المتاحة إلى فترتين: 1970 – 1986 و1982 – 1995، أي قبل وبعد 1982.

الآن لدينا ثلاثة نماذج انحدار محتملة:

الفترة الزمنية 1970 – 1981

$$Y_t = \lambda_1 + \lambda_2 X_t + u_{1t} \qquad n_1 = 12 \qquad \textbf{(1.8.8)}$$

الفترة الزمنية 1982 – 1995

$$Y_t = \lambda_1 + \lambda_2 X_t + u_{2t} \qquad n_2 = 14 \qquad \textbf{(2.8.8)}$$

الفترة الزمنية 1970 – 1995

$$Y_t = \alpha_1 + \alpha_2 X_t + u_{1t} \qquad n_1 = (n_1 + n_2) = 26 \qquad \textbf{(3.8.8)}$$

انحدار (3.8.8) يفترض أنه لا يوجد فرق بين الفترتين الزمنيتين، وبالتالي يقدر العلاقة بين الادخار و DPI للفترة الزمنية كلها مكونة من 26 مشاهدة، بمعنى آخر، هذا الانحدار يفترض أن الجزء المقطوع من المحور الصادي، وكذلك معامل الميل يظلان كما هما خلال الفترة الزمنية كلها، بالتالي لا يوجد تغير هيكلي.

إذا كان هذا هو واقع الحال، فإن $\gamma_1 = \lambda_1 = \alpha_1$ و $\gamma_2 = \lambda_2 = \alpha_2$ نموذجان للانحدار (1.8.8) و(2.8.8) يفترضان أن الانحدار يختلف باختلاف الفترة الزمنية، أي أن الجزء

المقطوع من المحور الصادي والميل مختلفان، وهذا يتضح من المعاملات المعبر عنهما. في الانحدار السابق، u's تمثل مقدار الخطأ و n's تمثل عدد المشاهدات.

باستخدام البيانات المعطاة في جدول (9.8)، النظير التطبيقي للانحدارات الثلاثة السابقة هي كالتالي:

$$\begin{aligned} \hat{Y}_t &= 1.0161 + 0.0803\,X_t \\ t &= (0.0873) \quad (9.6015) \\ R^2 = 0.9021 &\qquad \text{RSS}_1 = 1785.032 \qquad \text{df} = 10 \end{aligned} \tag{1a.8.8}$$

$$\begin{aligned} \hat{Y}_t &= 153.4947 + 0.0148X_t \\ t &= (4.6922) \quad (1.7707) \\ R^2 = 0.2971 &\qquad \text{RSS}_2 = 10{,}005.22 \qquad \text{df} = 12 \end{aligned} \tag{2a.8.8}$$

$$\begin{aligned} \hat{Y}_t &= 62.4226 + 0.0376\,X_t + \cdots \\ t &= (4.8917) \quad (8.8937) + \cdots \\ R^2 = 0.7672 &\qquad \text{RSS}_3 = 23{,}248.30 \qquad \text{df} = 24 \end{aligned} \tag{3a.8.8}$$

في الانحدارات السابقة، RSS تمثل مجموع مربعات البواقي، والأرقام الموضحة بين الأقواس هي قيم t المقدرة.

وبالنظر إلى مقدرات الانحدار، نجد أن العلاقة بين الادخار وDPI ليست ثابتة في الفترتين الزمنيتين. الميل في انحدارات الدخل - الادخار السابقة يمثل التغير الحدي للادخار (MPS)، وهنا هو متوسط التغير في الادخار كنتيجة لزيادة دولار واحد في الدخل.

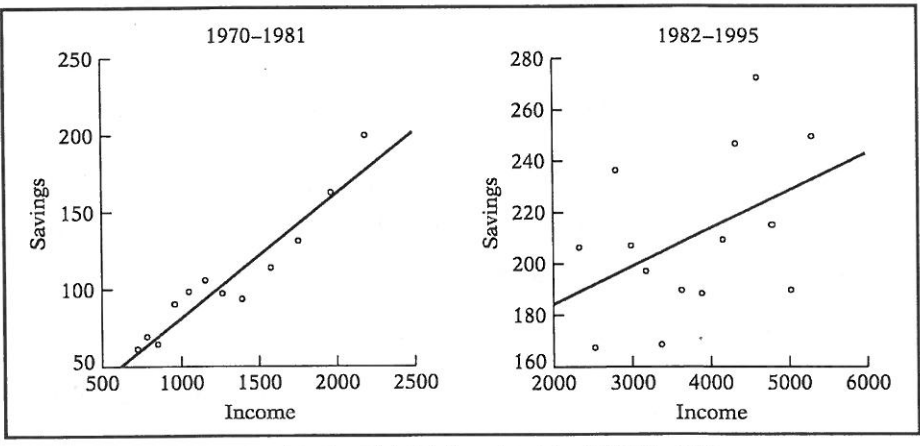
في الفترة 1970 - 1981، MPS كانت تقريباً 0.08، بينما تساوي 0.02 في الفترة 1982 - 1995. ومن الصعب تحديد ما إذا كان هذا التغير راجعًا للسياسات الاقتصادية التي اتخذها الرئيس ريجان أم لا، وهنا يقترح أن الانحدار المجمع (3a.8.8) الذي يتم استخدام الـ 26 مشاهدة مرة واحدة يتجاهل التغيرات المحتملة في الفترتين الزمنيتين، وهذا يعتبر شيئًا غير سليم أو غير مقبول.

بالطبع لإثبات العبارة السابقة، لابد من إجراء بعض الاختبارات الإحصائية، وبالأخص رسوم التشتت، ومعادلات الانحدار المقدرة، كما هو موضح في الشكل 3.8.

والآن الاختلافات المحتملة، والتي يمكن تسميتها بالتغيير الهيكلي، قد تحدث للجزء المقطوع من المحور الصادي أو للميل أو كليهما معاً.

كيف يمكن لنا تحديد ذلك؟ يمكن الاستعانة بالشكل (2.8) للإجابة على هذا السؤال، ولكن من المفيد أيضاً استخدام اختبارات إحصائية أخرى لذلك. وهنا يأتي الحديث عن اختبار chow [15]، وهذا الاختبار يفترض التالي:

1 - $u_{1t} \sim N(0, \sigma^2)$ ، $u_{2t} \sim N(0, \sigma^2)$. أي أن مقادير الأخطاء الخاصة بنماذج الانحدار تتبع التوزيع الطبيعي بنفس التباين σ^2 (ثبات التباين).



شكل (3.8) اختبارات إحصائية لمعادلات انحدار مقدرة

2 - مقدار الخطأ u_{1t} و u_{2t} مستقلان.

يتم اختبار Chow وفقاً للآلية الآتية:

1 - قدر انحدار (3.8.8)، وهو الذي يفترض عدم وجود استعراض المعلمات، واحصل على RSS_3 بدرجات حرية تساوي $n_1 + n_2 - k$ حيث k هي عدد المعلمات المقدرة وتساوي 2 في الحالة الراهنة محل الدراسة. في مثالنا الحالي، RSS_3=23,248.30. نحن نسمي RSS_3 مجموع مربعات البواقي المقيدة (RSS_R) حيث إنه يتم الحصول عليه وفقاً لمجموعة من القيود والخاصة بأن $\lambda_1 = \gamma_1$ و $\lambda_2 = \gamma_2$ مما يعني أن معاملات الانحدار لا تتغير.

(15) Gregory C. Chow, "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," Econometrica, vol. 28, no. 3, 1960, pp. 591–605.

2 - قدر (1.8.8) واحصل على مجموع مربعات البواقي RSS_1 بدرجات حرية $(n_1 - k)$. في مثالنا الحالي، $RSS_1 = 1785.032$ بدرجات حرية تساوي 10 .

3 - قدر (2.8.8) واحصل على مجموع مربعات البواقي RSS_2 بدرجات حرية $n_2 - k$. في مثالنا الحالي، $RSS_2 = 10{,}005.22$ بدرجات حرية = 12 .

4 - بما أن العينتين يفترض استقلالهما، يمكن أن نقوم بإضافة RSS_1 إلى RSS_2 ونحصل على ما يسمى مجموع مربعات الأخطاء غير المقيدة RSS_{UR} أي نحصل على التالي :

$$RSS_{UR} = RSS_1 + RSS_2$$

بدرجات حرية تساوي $= n_1 + n_2 - 2k$

في مثالنا الحالي :

$$RSS_{UR} = (1785.032 + 10{,}005.22) = 11{,}790.252$$

5 - والآن الفكرة من وراء اختبار chow، أنه إذا كان بالفعل لا يوجد تغير هيكلي [أي أن الانحدارين (1.8.8) و(2.8.8) بالضرورة متساويان] فإن RSS_R وRSS_{UR} لابد ألا يكون بينهما أي اختلاف إحصائي. وبالتالي إذا حصلنا على النسبة التالية :

$$F = \frac{(RSS_R - RSS_{UR})/k}{(RSS_{UR})/(n_1 + n_2 - 2k)} \sim F_{[k,(n_1+n_2-2k)]} \qquad (4.8.8)$$

يكون اختبار chow أوضح أنه تحت صحة الفرض العدمي الانحدارين (1.8.8) و(2.8.8) متساويان إحصائياً (بمعنى أنه لا يوجد تغيير أو انكسار هيكلي) والنسبة F المعطاة أسفل تتبع توزيع F بدرجات حرية K في البسط و$(n_1 + n_2 - 2k)$ في المقام.

6 - وبالتالي لا تستطيع رفض الفرض العدمي الخاص باستقرار المعلمات، حيث إن قيمة F المحسوبة عند تطبيق لا تتعدى قيمة F الحرجة التي نحصل عليها من جدول F عند مستوى المعنوية المحدد (أو قيمة p-value).

في مثل هذه الحالة، يمكن لنا أن نستخدم الانحدار المجمع (أو المقيد). وعلى العكس، إذا كانت قيمة F المحسوبة تزيد عن F الحرجة، فإننا نرفض الفرض عن انحدار (2.8.8). وفي هذه الحالة الأخيرة، يكون من غير السليم استخدام الانحدار المجمع وأقل ما يمكن قوله في هذه الحالة، أن قيمة مثل هذا الانحدار لا يمكن أن يعتد بها.

بالرجوع إلى مثالنا الحالي ، نجد أن :

$$F = \frac{(23,248.30 - 11,790.252)/2}{(11,790.252)/22}$$
$$= 10.69 \quad (5.8.8)$$

من جداول F ، نجد أنه لدرجات الحرية 2 و22 عند مستوى معنوية 1% قيمة F الحرجة تساوي 5.72. وبالتالي احتمال الحلول على قيمة لـ F مساوية أو أكبر من 10.69 أصغر بكثير من 1%، وللدقة أكثر، فإن p-value تساوي 0.00057 .

مما سبق، يتضح أن اختبار chow يدعم ما تم تصوره من قبل عن العلاقة بين الدخل - الادخار، وما تتضمن من تغيير هيكلي في الولايات المتحدة خلال الفترة 1970 - 1995، بافتراض صحة كل الفروض الخاصة بالاختبار. وسنناقش هذه النقطة بتفاصيل أكثر لاحقاً.

ويمكن ملاحظة أن اختبار chow يمكن تعميمه بسهولة للتعامل مع الحالات التي يكون فيها أكثر من تغيير هيكلي واحد. على سبيل المثال، إذا افترضنا أن العلاقة بين الدخل والادخار حدث فيها تغيير بعد تولي الرئيس كلينتون منصبه في يناير 1992، فمن الممكن تقسيم العينة إلى ثلاث مراحل : 1970 - 1981، 1982 - 1991، 1992 - 1995، وتستخدم اختبار chow بعد ذلك.

بالطبع سيكون لدينا أربعة RSS، قيمة واحدة لكل مرحلة جزيئية، وقيمة أخرى للبيانات المجمعة. ولكن المنطق وراء الاختبار يظل كما هو في البيانات حتى 2001 متاحة الآن لزيادة فترة الدراسة حتى نصل إلى عام 2001.

هناك بعض النقاط التي يجب وضعها في الاعتبار والمتعلقة باختبار chow وهي :

1 - يجب أن تتحقق كل الفروض الخاصة بهذا الاختبار. فعلى سبيل المثال، يجب على الباحث أن يتأكد من أن تباين الأخطاء في الانحدار (1.8.8)، والأخرى الخاصة بالانحدار (2.8.8) متساويان. وسنناقش هذه النقطة لاحقاً.

2 - اختبار chow سيحدد ما إذا كان الانحداران (1.8.8) و(2.8.8) مختلفين أم لا، ولكن بدون أي إشارة إلى طبيعة هذا الاختلاف فلن نعرف ما إذا كان ذلك راجعًا إلى الجزء المقطوع من المحور الصادي، أم راجع إلى الميل أم كليهما معاً. لكن في

الفصل (9) والخاص بالمتغيرات الوهمية، سنرى كيف يمكن الإجابة على هذا السؤال.

3 - اختبار chow يفترض معرفتنا بنقطة الانكسار الهيكلي في مثالنا الحالي، افترضنا أنها 1982. عموماً، إذا لم يكن من الممكن تحديد النقطة الزمنية التي حدث فيها التغيير الهيكلي، فيجب أن نستخدم طرقًا أخرى[16].

قبل أن نترك الحديث عن اختبار chow وانحدار (الدخل - الادخار) دعنا نختبر الفروض الخاصة باختبار chow، وبالأخص تباين الأخطاء الذي يعترض تساويه في الفترتين الزمنيتين.

حيث إننا لا نستطيع معرفة القيمة الحقيقية لتباين الأخطاء، يمكن أن نحصل على مقدراتهم من الـ RSS المعطى في انحدار (1a.8.8) وانحدار (2a.8.8)، كالتالي :

$$\hat{\sigma}_1^2 = \frac{\text{RSS}_1}{n_1 - 2} = \frac{1785.032}{10} = 178.5032 \qquad (6.8.8)$$

$$\hat{\sigma}_2^2 = \frac{\text{RSS}_2}{n_2 - 2} = \frac{10,005.22}{14 - 2} = 833.7683 \qquad (7.8.8)$$

لاحظ أنه، حيث إن لدينا معلمتين مقدرتين في كل معادلة، فإننا نطرح 2 من عدد المشاهدات للحصول على درجات الحرية، وتحت صحة الفروض الخاصة باختبار chow فإن $\hat{\sigma}_1^2$ ، $\hat{\sigma}_2^2$ مقدران غير متحيزين للتباين الحقيقي في الفترتين الزمنيتين.

وكنتيجة لذلك، يمكن ملاحظة أنه إذا كان $\sigma_1^2 = \sigma_2^2$ أي أن التباينات في الفترتين متساوية (كما يفترض اختبار chow) فإنه يمكن ملاحظة أن :

$$\frac{(\hat{\sigma}_1^2/\sigma_1^2)}{(\hat{\sigma}_1^2/\sigma_2^2)} \sim F_{(n_1-k),(n_2-k)} \qquad (8.8.8)$$

يتبع توزيع F بدرجات حرية $(n_1 - k)$ في البسط و$(n_2 - k)$ في المقام. في مثالنا الحالي، $k = 2$ بما أنه توجد معلمتان في كل انحدار جزيئي.

بالطبع تحت صحة أن $\sigma_1^2 = \sigma_2^2$ ، قيمة اختبار F السابقة يمكن اختصارها كالتالي :

$$F = \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} \qquad (9.8.8)$$

(16) لمناقشة هذه النقطة بالتفصيل، انظر William H. Greene, Econometric Analysis, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp. 293–297.

لاحظ أنه دائماً ما نضع القيمة الأكبر لمقدر التباين في البسط (انظر **ملحق A** للتفاصيل الخاصة بـ F، والتوزيعات الاحتمالية الأخرى) باستخدام قيمة F التي تم حسابها عن التطبيق، وقيمة F الحرجة بدرجات الحرية المناسبة، يمكن للفرد أن يحدد رفض أو عدم رفض الفرض العدمي والخاص بأن التباينين في المجتمعين الجزيئين متساويان إذا تم عدم رفض الفرض العدمي، فإنه يمكن استخدام اختبار chow في هذه الحالة.

بالرجوع إلى انحدارنا الخاص بالدخل والادخار، نحصل على النتيجة التالية:

$$F = \frac{833.7683}{178.5032} = 4.6701 \qquad \textbf{(10.8.8)}$$

تحت صحة الفرض العدمي الخاص بتساوي التباين في المجتمعين الجزيئين قيمة F تتبع توزيع F بدرجات حرية 12 في البسط و10 في المقام (لاحظ أننا وضعنا التباين المقدر الأكبر في البسط). من جداول F في **ملحق D**، نجد أن القيمة الحرجة للـ F عند 5 أو 1% لدرجات الحرية 12 و 10 هي على الترتيب 2.91 و 4.71. قيمة F المحسوبة معنوية عن 5% وتقريباً معنوية عن مستوى المعنوية 1%، وبالتالي فإنه يمكن استنتاج أن المجتمعين الجزيئين غير متساويين، وبالتالي أنه يمكن القول بشكل أكثر تحديداً أنه يمكن استخدام اختبار chow.

الهدف من هذه الفقرة، هو استعراض الآلية التي يمكن بها استخدام اختبار chow في التطبيقات الفعلية. إذا كانت تباينات الأخطاء في المجتمعين الجزيئين غير متساوية، فإنه يمكن تعديل اختبار chow. ولكن التفاصيل الخاصة بذلك خارج نطاق هذا الكتاب[17].

إذا كنا لا نستطيع تحديد النقطة الزمنية التي يحدث فيها الانكسار للعلاقة محل الدراسة، يمكن استخدام طرق بديلة مثل اختبار البواقي المعاد. سنستعرض هذه النقطة بالتفصيل في الفصل (13)، وهو الفصل الخاص بتحليل تحديد النموذج.

(17) لمناقشة تعديل اختبار chow لاستخدامه في حالات عدم تساوي أو ثبات التباين، انظر:
William H. Greene, Econometric Analysis, 4th ed., Prentice Hall, Englewood Cliffs, N.J., 2000, pp. 292–293, and Adrian C. Darnell, A Dictionary of Econometrics, Edward Elgar, U. K., 1994, p. 51.

9.8 التنبؤ باستخدام الانحدار المتعدد : PREDICTION WITH MULTIPLE REGRESSION

في الفقرة 10.5 أوضحنا أن نموذج الانحدار المقدر، والذي يحتوي على متغيرين اثنين يمكن أن يستخدم (1) للتنبؤ بالمتوسط، أي التنبؤ بالنقطة الموجودة على معادلة انحدار المجتمع (RPF) بالإضافة إلى (2) التنبؤ الفردي أي التنبؤ بقيمة فردية لـ Y عند أحد قيم المتغير المفسر X المعطاة $X = X_0$، حيث X_0 هي قيمة عددية محددة للمتغير X.

الانحدار المتعدد المقدر يمكن أن يستخدم أيضاً لأغراض مماثلة. والطرق الخاصة بإجراء ذلك هي امتداد تقليدي للحالة المقتصرة على متغيرين اثنين فقط، ما عدا أن المعادلات الخاصة بتقدير التباين والأخطاء المعيارية والمستخدمة في التنبؤ [مقارنة بـ (2.10.5) و(6.10.5) للنموذج ثنائي المتغيرات] من الأفضل والأسهل التعامل معها في صورة مصفوفات، وذلك مشروع في الملحق C. بالطبع معظم الأحزمة الإحصائية الخاصة بالانحدار، يمكن أن تستخدم للقيام بذلك، مما يجعل الاهتمام برموز المصفوفات غير ضروري. على الرغم من ذلك، فإن طرق المصفوفات معطاة في ملحق C للأهمية الرياضية لبعض الطلاب. هذا الملحق يعطي أيضاً أمثلة كاملة الحل والشرح باستخدام أسلوب المصفوفات.

10.8 (*) اختبارات الفروض الثلاثية: نسبة الإمكان (LR)، والـ wald tests (W)، معامل لاجرنش largrange (LM) THE TROIKA OF HYPOTHESIS TESTS: THE LIKELIHOOD RATIO (LR), WALD (W), AND LAGRANGE MULTIPLIER (LM) TESTS[18]:

في هذا الفصل والفصول السابقة، استخدمنا بشكل كبير اختبارات t، F وكاي - التربيعي، وذلك لإجراء العديد من الاختبارات في إطار نماذج الانحدار الخطية (في المعالم). ولكن بمجرد البعد عن نطاق نماذج الانحدار الخطية، نحتاج إلى طرق لاختبار فروض يمكن أن تتعامل مع نماذج خطية أو غير خطية.

(*) اختياري.

(18) لمزيد من التفاصيل، انظر: A. Buse, "The Likelihood Ratio, Wald and Lagrange Multiplier Tests: An Expository Note," American statistician, vol. 36, 1982, pp. 153–157.

الثلاثي المعروف والخاص بالإمكان و Wald ومعامل Lagrange يمكن استخدامه في هذا الفرض. والشيء المثير للإعجاب هو ملاحظة أن كل هذه الاختبارات تؤول تقريباً إلى توزيع كاي - التربيعي (أي في أحجام العينات الكبيرة نسبياً).

وعلى الرغم من أننا ناقشنا اختبار نسبة الإمكان في الملحق الخاص بهذا الفصل، ففي العموم لن نستخدم هذه الاختبارات في هذا الكتاب لاعتياد أحجام العينات الصغيرة أو المحدودة ذلك الاعتياد الذي للأسف يتعرض له معظم الباحثين، اختبار F والذي استخدمناه حتى الآن سيكون كافياً. فكما لاحظ Davidson و Mackinnon :

لنماذج الانحدار الخطية، سواء كانت الأخطاء تتبع أو لا تتبع التوزيع الطبيعي، لا توجد أي ضرورة للتعامل مع LU، W أو LR على الإطلاق، حيث إنه لا توجد أي فائدة من القيام بذلك تزيد عن الفائدة التي نحصل عليها عندما نستخدم اختبار F. [19]

11.8[*] اختبار الشكل الدالي للانحدار.. الاختيار بين نماذج الانحدار الخطية والخطية اللوغاريتمية :

TESTING THE FUNCTIONAL FORM OF REGRESSION: CHOSSING BETWEEN LINEAR AND LOG-LINEAR REGRESSION MODELS:

الاختيار بين نموذج الانحدار الخطي (المتغير المنحدر دالة خطية في المتغيرات المنحدر عليها) أو نموذج الانحدار الخطي اللوغاريتمي (لوغاريتم المتغير المنحدر دالة في لوغاريتم المتغيرات المنحدر عليها) هو سؤال مهم في التحليل التطبيقي. يمكن أن نستخدم اختبار Mackinnon، which و Davidson والذي نطلق عليه اختصاراً اختبار **MWD** والذي يمكننا من الاختيار بين النموذجين[20].

لشرح هذا الاختيار دعنا نفترض التالي:

H_0 : النموذج الخطي: Y دالة خطية في المتغيرات X's.

(*) اختياري.

(19) Russell Davidson and James G. MacKinnon, Estimation and Inference in Econometrics, Oxford University Press, New York, 1993, p. 456.

(20) J. MacKinnon, H. White, and R. Davidson, "Tests for Model Specification in the Presence of Alternative Hypothesis; Some Further Results." Journal of Econometrics, vol. 21, 1983, pp. 53–70. A similar test is proposed in A. K. Bera and C. M. Jarque, "Model Specification Tests: A Simultaneous Approach," Journal of Econometrics, vol. 20, 1982, pp. 59–82.

H_1: النموذج الخطي اللوغاريتمي : لوغاريتم Y دالة خطية في لوغاريتم الـ X's .

حيث كالمعتاد H_0 ترمز للفرض العدمي، و H_1 ترمز للفرض البديل .

اختبار MWD يقوم على الخطوات التالية :[21]

الخطوة I : قدر النموذج الخطي واحصل على قيم Y المقدرة، وتسمى هذه القيم Yf (أي $\hat{Y}$).

الخطوة II : قدر النموذج الخطي اللوغاريتمي واحصل على قيم $\ln Y$ المقدرة، وتسمى هذه القيم $\ln f$ (أي $\widehat{\ln Y}$).

الخطوة III : احصل على Z_1 والتي تساوي $\ln f - \ln Yf -$

الخطوة IVI : قم بعمل انحدار لـ Y على X's و Z_1 التي حصل عليها في الخطوة III . ارفض الفرض العدمي إذا كان معامل Z_1 معنويًا إحصائياً باستخدام إحضار t المعتاد .

الخطوة V : احصل على Z_2 والتي تساوي اللوغاريتم العكسي لـ $Yf - \ln f$.

الخطوة IV : قم بعمل انحدار لوغاريتم Y على لوغاريتم X's و Z_2 ارفض الفرض البديل إذا كان معامل Z_2 معنويًا إحصائياً باستخدام اختبار t المعتاد .

على الرغم من الصعوبة النسبية لاختبار MWP ، إلا أن المنطق من ورائه بسيط نسبياً. فإذا كان النموذج الخطي هو فعلاً النموذج السليم، فإن المتغير المكون Z_1 يجب ألا يكون معنويًا إحصائياً في الخطوة VI حيث إن القيم المقدرة الـ Y من النموذج الخطي، والأخرى المقدرة من النموذج الخطي اللوغاريتمي ستكون غير مختلفة . (لاحظ أننا نحصل على اللوغاريتم العكسي حتى يمكن أن نقارن بين القيم). نفس التعليق يمكن تطبيقه في حالة الفرض العدمي H_1.

---

(21) هذه المناقشة معتمدة على المرجع التالي :
William H. Greene, ET. The Econometrics Toolkit Version 3, Econometric Software, Bellport, New York, 1992, pp. 245–246.

مثال 5.8

الطلب على الورد : The Demand for roses

وبالرجوع إلى تمرين 16.7 والذي يحتوي على بيانات خاصة بالطلب على الورد في المنطقة السكنية المحيطة بمدن De troit في الفترة 11-1971 إلى 11-1975 . وللتوضيح وللتبسيط سنفرض أن الطلب على الزهور هو دالة في أسعار الورد وأسعار القرنفل ولن تستخدم متغير الدخل في الوقت الحالي . الآن دعنا نعرض النماذج التالية :

(1.11.8) النموذج الخطي : $Y_t = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + u_i$

(2.11.8) النموذج الخطي اللوغاريتمي : $\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + u_i$

حيث : Y هي الكمية التامة من الورد بالدستة ، X_2 هو متوسط سعر الورد (دولار/ الدستة)، X_3 متوسط سعر القرنفل (دولار/ الدستة) . مسبقاً المفروض أن α_2 وβ_2 متوقع أن يحمل الإشارة السالبة (لماذا؟) . وα_2 وβ_3 يحملان الإشارة الموجبة (لماذا؟) . كما نعلم فإن معامل الميل في النموذج الخطي اللوغاريتمي يمثل معامل المرونة .

نتائج الانحدار كالتالي :

$$\begin{aligned} \hat{Y}_t &= 9734.2176 - 3782.1956 X_{2t} + 2815.2515 X_{3t} \\ t &= (3.3705) \quad (-6.6069) \quad (2.9712) \\ F &= 21.84 \qquad R^2 = 0.77096 \end{aligned} \tag{3.11.8}$$

$$\begin{aligned} \widehat{\ln Y_t} &= 9.2278 - 1.7607 \ln X_{2t} + 1.3398 \ln X_{3t} \\ t &= (16.2349) \quad (-5.9044) \quad (2.5407) \\ F &= 17.50 \qquad R^2 = 0.7292 \end{aligned} \tag{4.11.8}$$

كما توضح هذه النتائج ، فإن كلاً من النموذج الخطي ، والنموذج الخطي اللوغاريتمي مناسبان بشكل جيد للبيانات محل الدراسة ، حيث إن المعالم لها الإشارات المتوقعة وقيم t وR^2 معنوية إحصائياً .

للاختيار بين هذه النماذج على أساس اختبار **MWD**، نختبر أولاً الفرض الخاص بأن النموذج السليم هو النموذج الخطي . ثم باتباع الخطوة VI للاختبار ، نحصل على الانحدار التالي :

$$\begin{aligned} \hat{Y}_t &= 9727.5685 - 3783.0623 X_{2t} + 2817.7157 X_{3t} + 85.2319 \\ t &= (3.2178) \quad (-6.3337) \quad (2.8366) \quad (0.0207) \\ F &= 13.44 \qquad R^2 = 0.7707 \end{aligned} \tag{5.11.8}$$

وحيث إن معامل Z_1 ليس معنويًا إحصائياً (قيمة p-value للـ t المقدرة هي 0.98) فإننا لا نستطيع رفض الفرض القائل بأن النموذج السليم هو النموذج الخطي .

دعنا نفترض أننا بدلنا الفرض بأن النموذج السليم هو النموذج الخطي اللوغاريتمي باتباع الخطوة IV من خطوات الاختبار MWP، نحصل على نتائج الانحدار التالية :

$$\widehat{\ln Y_t} = 9.1486 - 1.9699 \ln X_t + 1.5891 \ln X_{2t} - 0.0013 Z_{2t}$$
$$t = (17.0825) \quad (-6.4189) \quad (3.0728) \quad (-1.6612) \qquad \textbf{(6.11.8)}$$
$$F = 14.17 \qquad R^2 = 0.7798$$

وبما أن معامل Z_2 معنوي إحصائياً عند مستوى معنوية 12 (p-value تساوي 0.1225) بالتالي، فإننا نرفض الفرض العدمي الخاص بأن النموذج السليم هو النموذج الخطي اللوغاريتمي عند هذا المستوى من المعنوية. بالطبع إذا اتبعنا مستوى المعنوية المعتاد 1% أو 5% فإننا لن تستطيع رفض الفرض الخاص بأن النموذج السليم هو النموذج الخطي اللوغاريتمي. وكما يتضح من المثال الحالي، فإننا في بعض الحالات لا نستطيع رفض أي من النموذجين.

12.8 الخلاصة والاستنتاج : SUMMARY AND CONCLUSIONS

1 - هذا الفصل يستعرض ويضيف أبعاداً جديدة للأفكار الخاصة بالتقدير مرة، واختبارات الفروض التي تم استعراضها من قبل في الفصل (5) في إطار نماذج الانحدار الخطية ثنائية المتغيرات.

2 - في الانحدار المتعدد، اختيار معنوية أحد معاملات الانحدار منفرداً (باستخدام اختبار t) واختبار معنوية العلاقة ككل (أي يكون الفرض العدمي H_0 : كل معاملات الانحدار تساوي الصفر أو $R^2 = 0$) يعتبران اختبارين مختلفين تماماً.

3 - إذا أردنا التحدث بشكل أكثر دقة، فإنه إذا توصلنا إلى أن واحداً أو أكثر من معاملات الانحدار الجزيئية معنوي إحصائياً على أساس اختبار t المنفرد، فإن ذلك لا يعني بالضرورة أن كل معاملات الانحدار الجزيئية معنوية بشكل (تجميعي) إحصائياً. وهذا الفرض الأخير من الممكن اختباره باستخدام اختبار F.

4 - اختبار F متفوق في هذا المجال، حيث إنه يسمح بعمل العديد من اختبارات الفروض، مثل التالي (1) اختبار معنوية معامل الانحدار منفرداً، (2) كل معاملات الميل الجزيئية تساوي الصفر، (3) اثنان أو أكثر من المعاملات متساويان إحصائياً) (4) المعاملات تحقق بعضاً من القيود الخطية و(5) اختبار وجود استقرار هيكلي في نموذج الانحدار.

5 - كما في حالة متغيرين اثنين، الانحدار المتعدد ممكن استخدامه بفرض التنبؤ بالمفردات أو المتوسط.

تمــاريــن : EXERCISES

1.8 افترض أنك تريد دراسة سلوك مبيعات سلعة ما، مثلاً، السيارات خلال فترة زمنية من السنوات، وافترض أن أحدًا اقترح عليك أن تجرى النماذج التالية :

$$Y_t = \beta_0 + \beta_1 t$$

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2$$

حيث : Y_t = المبيعات عند الزمن t، t = الزمن مقاس بالسنوات. النموذج الأول يفترض أن المبيعات هي دالة خطية في الزمن، في حين النموذج الثاني يفترض أنها دالة تربيعية في الزمن.

(a) ناقش خصائص هذه النموذج.

(b) كيف يمكن أن تختار بين هذين النموذجين؟

(c) في أي حالات يمكن أن يكون النموذج التربيعي مناسبًا؟

(d) حاول أن تحصل على بيانات عن مبيعات السيارات في الولايات المتحدة في العشرين سنة السابقة، وحدد أيًا من النموذجين مناسب أكثر للبيانات.

2.8 اثبت أن النسبة F الموجودة في (16.5.8) تساوي النسبة F الموجودة في (18.5.8) (لاحظ أن : $ESS/TSS = R^2$).

3.8 اثبت أن اختبارات F الموجودة في (18.5.8) و (10.7.8) متساويان.

4.8 تحقق من العلاقات الموجودة في (11.7.8) و (12.7.8).

5.8 اعتبر دالة إنتاج Cobb - Douglas

$$Y = \beta_1 \; L^{\beta_2} \; K^{\beta_3} \quad (1)$$

حيث : Y = الناتج، L = مدخل العمالة، K = مدخل رأس المال بقسمة (1) على K نحصل على

$$(Y/K) = \beta_1 \; (L/K)^{\beta_2} \; K^{\beta_3+\beta_2-1} \quad (2)$$

وبإدخال اللوغاريتم الطبيعي على الطرفين، وإضافة مقدار الخطأ، نحصل على

$$\ln (Y/K) = \beta_0 + \beta_2 \ln (L/K) + (\beta_2 + \beta_3 - 1) \ln K + u_i \quad (3)$$

حيث : $\beta_0 = \ln \beta_1$

(a) افترض أن لديك بيانات و استخدمت معها النموذج (3) .

كيف يمكنك اختبار ثبات العائد على المقياس . أي $(\beta_2 + \beta_3 = 1)$؟

(b) إذا كان العائد ثابتًا ، كيف يمكن لك أن تفسر الانحدار؟

(c) هل هناك أي فرق من قسمة (1) على L بدلاً من K؟

6.8 القيم الحرجة للـ R^2 عندما تكون R^2 الحقيقية = الصفر .

المعادلة (11.5.8) تمثل العلاقة بين F ، R^2 تحت صحة الفرض الخاص بأن كل معاملات الميل الجزيئية بشكل متتابع تساوي الصفر . (أي أن $R^2 = 0$) . ومن خلال جدول F يمكننا إيجاد قيمة F الحرجة عند مستوى المعنوية α ، وبالتالي نحصل على قيمة R^2 الحرجة من العلاقة التالية :

$$R^2 = \frac{(k-1)F}{(k-1)F + (n-k)}$$

حيث : K هي عدد معاملات نموذج الانحدار مشتملة على الجزء المقطوع من المحور الصادي ، وF هي القيمة الحرجة للـ F عند مستوى المعنوية α . إذا كانت القيمة المشاهدة للـ R^2 تزيد عن R^2 الحرجة التي نحصل عليها من العلاقة السابقة ، فإننا نرفض الفرض الخاص بأن R^2 الحقيقية تساوي الصفر .

تحقق من العلاقة السابقة ، وأوجد قيم R^2 الحرجة (عند مستوى معنوية α = 5%) وذلك للانحدار الموجود في (1.2.8) .

7.8 باستخدام بيانات سنوية للفترة 1968 – 1987 تم الحصول على نتائج الانحدار التالية :

$$\hat{Y}_t = -859.92 + 0.6470\, X_{2t} = 23.195\, X_{3t} \qquad R^2 = 0.9776 \qquad (1)$$

$$\hat{Y}_t = -261.09 + 0.2452\, X_{2t} \qquad R^2 = 0.9388 \qquad (2)$$

حيث : Y = إنفاق الولايات المتحدة على السلع المستوردة مقدرة بالبليون دولار سنة 1982 ، X_2 = الدخل الشخصي مقدر بالبليون دولار في 1982 ، X_3 = متغير الاتجاه .

حدد ما إذا كان التالي صح أم خطأ :

الخطأ المعياري لـ X_3 في (1) يساوي 4.2750 .

وضح الخطوات الحسابية لإجابتك . (ملاحظة : استخدم العلاقة بين R^2 ، F و t) .

8.8 افترض أن لدينا الانحدار التالي :

$$\ln (Y_i / X_{2i}) = \alpha_1 + \alpha_2 \ln X_{2i} + \alpha_3 \ln X_{3i} + u_i$$

افترض أن قيم معاملات الانحدار وأخطاءها المعيارية معروفة(*) .

من خلال معرفتنا بهذه القيم ، كيف يمكن تقدير معلمات نموذج الانحدار التالي وأخطاءها المعيارية؟

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i$$

9.8 افترض التالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} + u_i$$

حيث : Y = نفقات الاستهلاك الشخصي ، X_2 = الدخل الشخصي و X_3 = الثروة الشخصية(**) . المقدار $(X_{3i} X_{2i})$ معروف بحد التفاعل . ما هو المقصود بهذا التعبير؟ كيف يمكنك اختبار أن الكثافة الحدية للاستهلاك (MPC) (أي β_2) مستقلة عن ثروة المستهلك؟

10.8 افترض أن لديك نتائج الانحدار التالية :

$$\hat{Y}_t = 16{,}899 \quad - 2978.5 X_{2t} \qquad R^2 = 0.6149$$
$$t = \quad (8.5152) \quad (-4.7280)$$
$$\hat{Y}_t = 9734.2 \quad - 3782.2 X_{2t} \quad + \quad 2815 X_{3t} \qquad R^2 = 0.7706$$
$$t = \quad (3.3705) \quad (-6.6070) \quad (2.9712)$$

هل يمكنك إيجاد حجم العينة الخاصة بهذه النتائج؟

(ملاحظة : استخدم العلاقة بين قيم R^2 ، F و t) .

11.8 بناء على مناقشتنا السابقة عن اختبارات الفروض الفردية والمشتركة والخاصة باختبار t و F على الترتيب ، أي من المواقف التالية الأقرب للحدوث؟

1 - رفض الفرض المشترك باستخدام إحصاء F ، وعدم رفض كل فرض منفرد على أساس اختبارات t المنفردة .

(*) مقتبس من : Peter Kennedy, A Guide to Econometrics, the MIT Press, 3d ed., Cambridge, Mass., 1992, p. 310.

(**) Ibid., p. 327.

2 - رفض الفرض المشترك على أساس إحصاء F، ورفض اختبار واحد لمعلمة انحدار منفردة على أساس اختبار t، وعدم رفض باقي الاختبارات المنفردة للمعاملات الأخرى على أساس اختبار t.

3 - رفض الفرض المشترك على أساس إحصاء F، ورفض كل اختبار لمعنوية المعاملات منفردة على أساس اختبار t.

4 - عدم رفض الفرض المشترك على أساس إحصاء F، ورفض كلٍ من المعاملات منفرداً على أساس اختبارات t(*).

مسائل : Problems

12.8 بالإشارة إلى تمرين 21.7

(a) ما هي مرونة معدل الفائدة؟ ومرونة الدخل الحقيقي للرصيد النقدي الحقيقي؟

(b) هل المرونات السابقة لها معنوية إحصائياً كل منها على حدة؟

(c) اختبر معنوية العلاقة ككل للانحدار المقدر.

(d) هل مرونة الطلب على الرصيد النقدي الحقيقي تختلف معنوياً عن الواحد؟

(e) هل من الضروري الإبقاء على متغير معدل الفائدة في النموذج؟ ولماذا؟

13.8 من بيانات 46 ولاية في الولايات المتحدة عام 1992، حصل Baltagi على نتائج الانحدار التالية(**):

$$\widehat{\log C} = 4.30 - 1.34 \log P + 0.17 \log Y$$
$$\text{se} = (0.91) \quad (0.32) \qquad (0.20) \qquad \bar{R}^2 = 0.27$$

حيث : C = استهلاك السجائر، مقدر بعدد العبوات في العام الواحد.

P = السعر الحقيقي للعبوة.

Y = الدخل الحقيقي للفرد.

(a) ما هي مرونة الطلب على السجائر بالنسبة السعرها؟ هل لها معنوية إحصائية؟ وإذا كان ذلك صحيحًا، هل تختلف معنوياً عن الواحد؟

(b) ما هي مرونة الدخل بالنسبة للطلب على السجائر؟ هل لها معنوية إحصائية؟ وإذا كان ذلك غير صحيح، ما هي الأسباب الممكنة لذلك؟

(*) مقتبسة من : Ernst R. Berndt, The Practice of Econometrics: Classic and Contemporary, Addison-Wesley, Reading, Mass., 1991, p. 79.

(**) انظر : Badi H. Baltagi, Econometrics, Springer-Verlag, New York, 1998, p. 111.

(c) كيف يمكنك الحصول على R^2 من R^2 المعدلة والمعطاة سابقاً؟

14.8 بناء على عينة مكونة من 209 مشروعات، حصل Wooldridge على نتائج الانحدار التالية:(*)

$$\log(\widehat{salary}) = 4.32 + 0.280 \log(sales) + 0.0174 \, roe + 0.00024 \, ros$$
$$se = (0.32) \quad (0.035) \quad (0.0041) \quad (0.00054)$$
$$R^2 = 0.283$$

حيث: Salary = الراتب الخاص بـ CEO

Sales = المبيعات السنوية للمشروع

roe = عائد نسبة حقوق الملكية

ros = عائد مخزون المشروع

والأرقام داخل الأقواس تعبر عن الأخطاء المعيرية المقدرة.

(a) فسر نتائج الانحدار السابقة واضعاً في الاعتبار أي توقعات مسبقة عن إشارات المعاملات المختلفة.

(b) أي من المعاملات السابقة له معنوية إحصائية منفرداً عند مستوى معنوية 5%؟.

(c) ما هي المعنوية الكلية للانحدار؟ ما هو الاختبار المستخدم؟ ولماذا؟

(d) هل يمكن تفسير معامل roe و ros كمعاملات مرونة؟ علل إجابتك سواء كانت بنعم أو لا؟

15.8 افترض أن Y و X_2، X_3، . . . ، X_k لها توزيع طبيعي مشترك، وافترض أننا نريد اختبار الفرض العدمي القائل بأن معاملات الارتباط الجزئية في المجتمع تساوي الصفر كل على حدة، R.A.Fisher أثبت أن:

$$t = \frac{r_{1\,2.3\,4\ldots k}\sqrt{n-k-2}}{\sqrt{1-r^2_{1\,2.3\,4\ldots k}}}$$

تتبع توزيع t بدرجات حرية $(n-k-2)$، حيث k هي معامل الارتباط الجزئي من الدرجة k، و n هي عدد المشاهدات الكلي (لاحظ أن: $r_{1\,2.3}$ هو معامل ارتباط جزئي من الدرجة الأولى، $r_{1\,2.34}$ هو معامل ارتباط جزئي من الدرجة الثانية، وهكذا).

(*) انظر: Jeffrey M. Wooldridge, Introductory Econometrics, South-Western Publishing Co., 2000, pp. 154–155.

بالإشارة إلى تمرين 2.7 مفترضين أن Y و X_2 و X_3 لها توزيع طبيعي مشترك، احسب كلاً من معاملات الارتباط الجزيئية الثلاثة التالية: $r_{12.3}$، $r_{13.2}$، $r_{23.1}$ واختبر معنويتها تحت صحة الفرض القائل بأن معامل ارتباط المجتمع الخاص بكل حالة يساوي الصفر، كلاً على حدة.

16.8 في دراسة خاصة بالطلب على جرارات المزارع في الولايات المتحدة خلال الفترة 1921 – 1941 و 1948 – 1957، حصل Criliches(*) على النتائج التالية:

$$\widehat{\log Y_t} = \text{constant} - \underset{(0.231)}{0.519} \log X_{2t} - \underset{(0.477)}{4.933} \log X_{3t} \qquad R^2 = 0.793$$

حيث: Y_t = قيمة مخزون الجرارات في المزارع مقاس في 1 يناير عام 1935 إلى 1939 بالدولار، X_2 مؤشر الأسعار المدفوعة للجرارات مقسومًا على مؤشر أسعار المحاصيل الزراعية عند الزمن $t-1$، X_3 = معدل الفائدة في العام $t-1$ والأخطاء المعيارية المقدرة هي الأرقام المعطاة ما بين الأقواس!

(a) فسر الانحدار السابق.

(b) هل معاملات الميل المقدرة لها معنوية إحصائية كلاً على حدة؟ هل يختلفون عن الواحد الصحيح؟

(c) استخدم تحليل التباين لاختبار معنوية نموذج الانحدار ككل.

ملاحظة: استخدم R^2 المعطاة في أسلوب ANOVA.

(d) كيف يمكنك حساب مرونة معدل الفائدة للطلب على جرارات المزارع؟

(e) كيف يمكنك اختبار معنوية R^2 المقدرة؟

17.8 اعتبر معادلة تحديد الأجور التالية والخاصة بالاقتصاد البريطاني في الفترة 1950 – 1969.

$$\hat{W}_t = \underset{(1.129)}{8.582} + \underset{(0.080)}{0.364}(\text{PF})_t + \underset{(0.072)}{0.004}(\text{PF})_{t-1} - \underset{(0.658)}{2.560}U_t$$

$$R^2 = 0.873 \qquad \text{df} = 15$$

(*) Z. Griliches, "The Demand for a Durable Input: Farm Tractors in the United States, 1921–1957," in The Demand for Durable Goods, Arnold C. Harberger (ed.), The University of of Chicago Press, Chicago, 1960, Table 1, p. 192.

حيث : W = الأجور والرواتب محسوبة للفرد الواحد .

PF = أسعار المنتج النهائي عند تكلفة المصنع .

U = البطالة في بريطانيا العظمى مقدرة كنسبة من إجمالي العمالة فيها .

t = الزمن

(الأرقام ما بين الأقواس تمثل الأخطاء المعيارية المقدرة)

(a) فسر المعادلة السابقة .

(b) هل المعاملات المقدرة معنوية؟ كلاً على حدة .

(c) ما هو التبرير المنطقي لإدخال $(PF)_{t-1}$ للمعادلة؟

(d) هل يجب حذف المتغير $(PF)_{t-1}$ من النموذج؟ لماذا؟

(e) كيف يمكنك حساب مرونة الأجر والراتب للموظف الواحد بالنسبة لمعدل البطالة U؟

18.8 المعادلة التالية هي معادلة أخرى لتحديد الأجر مختلفة عن المعطاة في تمرين 17.8 كالتالي :

$$\hat{W}_t = 1.073 + 5.288V_t - 0.116X_t + 0.054M_t + 0.046M_{t-1}$$
$$(0.797) \quad (0.812) \quad (0.111) \quad (0.022) \quad (0.019)$$
$$R^2 = 0.934 \qquad df = 14$$

حيث : W = الأجر والراتب للموظف الواحد .

V = وظائف خالية لم يتم شغلها بعد في بريطانيا العظمى محسوبة كنسبة من عدد العمالة الكلية بها .

X = الإنتاج المحلي الكلي بالنسبة للشخص العامل الواحد .

M = أسعار الاستيراد .

M_{t-1} = أسعار الاستيراد في السنة السابقة .

(الأخطاء المعيارية المقدرة معطاة بين الأقواس) .

(a) فسر المعادلة السابقة .

(b) أي من المعاملات المقدرة له معنوية إحصائية؟ كلاً على حدة .

(c) ما هو التبرير المنطقي لإدخال المتغير X في النموذج؟ وهل يمكن التوقع مسبقاً بأن إشارة X سالبة؟

(d) ما هو سبب إدخال كل من M_t و M_{t-1} إلى النموذج؟

(e) أي من المتغيرات يمكن حذفه من النموذج؟ لماذا؟

(f) اختبر المعنوية الكلية للانحدار المقدر.

19.8 بالعودة إلى دالة الطلب على الدجاج المقدرة فى (24.7.8)، هل مرونة الدخل المقدرة تساوي 1؟ هل مرونة السعر تساوي 2- ؟

20.8 باستخدام دالة الطلب الموجودة في (24.7.8) كيف يمكنك اختبار الفرض القائل بأن مرونة الدخل تساوي كقيمة، ولكن بإشارة معكوسة مرونة السعر على الطلب؟ وضح الخطوات الحسابية

[لاحظ أن: $\text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.00142$]

21.8 بالإشارة إلى تمرين 16.7 والخاص بدالة الطلب على الورد. وباستخدام الدالة اللوغاريتمية أجب على التالي:

(a) ما هي المرونة السعرية المقدرة للطلب (أي المرونة بالنسبة لسعر الورد)؟

(b) هل لها معنوية إحصائية؟

(c) إذا كان ذلك صحيحًا، هل للمرونة معنوياً تختلف عن الواحد الصحيح؟

(d) مسبقاً، ما هي الإشارة المتوقعة لـ X_3 (سعر القرنفل) و X_4 (الدخل)؟ هل النتائج التطبيقية متماشية مع هذه التوقعات؟

(e) إذا كانت معاملات X_3 و X_4 غير معنوية إحصائياً، ما هي الأسباب المتوقعة لذلك؟

22.8 بالإشارة إلى تمرين 17.7 والخاص بإنشطة القطط الوحشية:

(a) هل لكل معامل ميل، كلاً على حدة، معنوية إحصائياً عند مستوى معنوية 5%.

(b) هل ترفض الفرض القائل أن $R^2 = 0$ ؟

(c) ما هو معدل النمو اللحظي لأنشطة القطط الوحشية خلال الفترة 1948 إلى 1978 ؟ وما هو معدل النمو المركب المرتبط به؟

23.8 بالإشارة إلى ميزانية الدفاع في الولايات المتحدة والموضحة في نموذج الانحدار المقدر في تمرين 18.7 أجب عن التالي:

(a) علق بشكل عام على نتائج الانحدار المقدر.

(b) كون جدول الـ ANOVA واختبر الفرض الخاص بأن كل معاملات الميل الجزيئية تساوي الصفر.

24.8 الدالة التالية تسمى دالة الإنتاج الفائقة (TPF) وهي تعبر عن تعميم للدالة المعروفة بدالة إنتاج Cobb-Dauglas:

$$Y_i = \beta_1 L^{\beta_2} k^{\beta_3} e^{\beta_4 L + \beta_5 K}$$

حيث: Y = الناتج، L = مدخل العمالة، K = مدخل رأس المال. باستخدام اللوغاريتم على طرفي الدالة، وإضافة مقدار خطأ عشوائي نحصل على TPF العشوائية كالتالي:

$$\ln Y_i = \beta_0 + \beta_2 \ln L_i + \beta_3 \ln K_i + \beta_4 L_i + \beta_5 K_i + u_i$$

حيث: $\beta_0 = \ln \beta_1$

(a) ما هي خصائص هذه الدالة؟

(b) لصياغة TPF في شكل مختصر مماثل لدالة إنتاج (Cobb-Douglas)، ما هي القيم اللازمة لكل من β_4 وβ_5؟

(c) إذا توافرت لدينا بيانات، كيف يمكنك معرفة ما إذا كان ممكنًا اختصار TPF في صورة دالة إنتاج Cobb-Douglas؟
ما هي خطوات الاختبار التي سنستخدمها؟

(d) اختبر ما إذا كان TPF تناسب بيانات الجدول (8.8)، وضح خطواتك الحسابية.

25.8 تكوين رأس المال وأسعار الطاقة: الولايات المتحدة، 1948 - 1978 لاختبار الفرض القائل بأن ارتفاع أسعار الطاقة بالنسبة للمنتجات يؤدي إلى انخفاض إنتاجية رأس المال الحالي والعمالة، قام John A. Taton بتقدير دالة الإنتاج التالية للولايات المتحدة للفترة الربع سنوية 1948–I إلى 1978–II(*):

$$\widehat{\ln(y/k)} = \underset{(16.33)}{1.5492} + \underset{(21.69)}{0.7135} \ln(h/k) - \underset{(-6.42)}{0.1081} \ln(P_e/P)$$

$$+ \underset{(15.86)}{0.0045t} \qquad R^2 = 0.98$$

(*) انظر التالي: Energy Prices and Capital Formation: 1972–1977," Review, Federal Reserve Bank of St. Louis, vol. 61, no. 5, May 1979, p. 4.

حيث : y = الناتج الحقيقي لقطاع الأعمال الخاص .
k = مقياس تدفق خدمات رأس المال .
h = ساعات الفرد في قطاع الأعمال الخاص .
P_e = مؤشر سعر المنتج للوقود والسلع المرتبطة به .
P = سعر قطاع الأعمال الخاص .
t = الزمن

الأرقام بين الأقواس تمثل قيم إحصاءات الـ t

(a) هل النتائج تؤيد فرض الباحث؟

(b) خلال الفترة 1972 إلى 1977 زاد السعر النسبي للطاقة (P_e/P) بحوالي 60% من الانحدار المقدر، ما هي الخسارة في الإنتاجية؟

(c) إذا سمحنا للقيم (h/k) و(P_e/P) للتغيير، ما هو اتجاه معدل النمو للإنتاجية خلال الفترة الزمنية السابقة؟

(d) كيف يمكنك تفسير قيمة المعامل 0.7135 ؟

(e) هل حقيقة أن معاملات الميل الجزيئية، كلاً على حدة، لها معنوية إحصائياً (فسر ذلك) تعني أننا نرفض الفرض الخاص بأن $R^2 = 0$ ؟ علل إجابتك .

26.8 الطلب على الأسلاك. جدول (10.8) يعطي البيانات الخاصة بمصنع أسلاك تليفونية. المطلوب التنبؤ بالمبيعات لعميل مهم ورئيس خلال الفترة 1968 إلى 1983 . (*)

المتغيرات في الجدول معرفة كالتالي:

Y = المبيعات السنوية مقدرة بالمليون قدم مربع MPF .
X_2 = الناتج القومي الكلي (GNP) مقدرة بالبليون دولار .
X_3 = عدد الوحدات المنزلية مقدرة بالآلاف .
X_4 = معدل البطالة ، % .
X_5 = المعدل الأولي في الستة أشهر السابقة .
X_6 = الزيادة في خطوط العملاء ، % .

(*) وافر الشكر إلى Daniel J. R لتجميع وتجهيز هذه البيانات .

Regnession Variables جدول (10.8) متغيرات الانحدار

Year	X_2, GNP	X_3, housing starts	X_4, unemployment, %	X_5, prime rate lag, 6 mos.	X_6, customer line gains, %	Y, total plastic purchases (MPF)
1968	1051.8	1503.6	3.6	5.8	5.9	5873
1969	1078.8	1486.7	3.5	6.7	4.5	7852
1970	1075.3	1434.8	5.0	8.4	4.2	8189
1971	1107.5	2035.6	6.0	6.2	4.2	7497
1972	1171.1	2360.8	5.6	5.4	4.9	8534
1973	1235.0	2043.9	4.9	5.9	5.0	8688
1974	1217.8	1331.9	5.6	9.4	4.1	7270
1975	1202.3	1160.0	8.5	9.4	3.4	5020
1976	1271.0	1535.0	7.7	7.2	4.2	6035
1977	1332.7	1961.8	7.0	6.6	4.5	7425
1978	1399.2	2009.3	6.0	7.6	3.9	9400
1979	1431.6	1721.9	6.0	10.6	4.4	9350
1980	1480.7	1298.0	7.2	14.9	3.9	6540
1981	1510.3	1100.0	7.6	16.6	3.1	7675
1982	1492.2	1039.0	9.2	17.5	0.6	7419
1983	1535.4	1200.0	8.8	16.0	1.5	7923

دعنا نعبر عن النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + \beta_5 X_{5t} + \beta_6 X_{6t} + u_t$$

(a) قدر الانحدار السابق.

(b) ما هي الإشارات المتوقعة لمعاملات النموذج السابق؟

(c) هل النتائج التطبيقية تتماشى مع التوقعات المسبقة؟

(d) هل معاملات الانحدار الجزيئية، كلاً على حدة، لها معنوية إحصائية عند مستوى المعنوية 5%؟

(e) افترض أنك قمت أولاً بعمل انحدار لـ Y على X_2، X_3، X_4 فقط، ثم قررت إضافة المتغيرات X_5 و X_6. كيف يمكنك معرفة أهمية إضافة هذين المتغيرين؟ ما هو الاختبار المستخدم لعمل ذلك؟ وضح الخطوات اللازمة لذلك.

27.8 قام Marc Nerlove بتقدير دالة التكلفة التالية والخاصة بمولدات الكهرباء كالتالي: (*)

$$R_{it} = \hat{\alpha}_i + \hat{\beta}_i R_{mt} + e_{it} \quad (1)$$

(*) Marc Nerlove, "Returns to Scale in Electric Supply," in Carl Christ, ed., Measurement in Economics, Stanford University Press, Palo Alto, Calif., 1963. The notation has been changed.

في هذه الفقرة تم تغيير الرموز عن الموجود في تلك الدراسة.

حيث : Y = التكلفة الكلية للإنتاج

X = الناتج مقدر بالكيلو - وات .

P_1 = سعر مدخل العمالة .

P_2 = سعر مدخل رأس المال .

P_3 = سعر الوقود .

u = مقدار الخطأ العشوائي .

نظرياً ، مجموع المرونات متوقع أن يساوي الواحد الصحيح ، أي أن $\alpha_1 + \alpha_2 + \alpha_3 = 1$. بناء على هذا الفرض ، فإن دالة التكلفة السابقة يمكن كتابتها كالتالي :

$$(Y/P_3) = AX^{\beta}(P_1/P_3)^{\alpha_1}(P_2/P_3)^{\alpha_2}u \qquad (2)$$

بمعني آخر ، فإن (1) هي صورة الدالة غير المقيدة ، أما (2) فهي دالة التكلفة المقيدة .

باستخدام عينة مكونة من 29 مشروعًا متوسط الحجم ، وبعد استخدام التحويلة اللوغاريتمية ، حصل Nerlove على نتائج الانحدار التالية :

$$\begin{aligned}\widehat{\ln Y_i} = &-4.93 + 0.94 \ln X_i + 0.31 \ln P_1 \\ \text{se} = &\ (1.96) \qquad (0.11) \qquad (0.23) \\ &-0.26 \ln P_2 + 0.44 \ln P_3 \\ &\ (0.29) \qquad (0.07) \qquad \text{RSS} = 0.336\end{aligned} \qquad (3)$$

$$\begin{aligned}\widehat{\ln (Y/P_3)} = &-6.55 + 0.91 \ln X + 0.51 \ln (P_1/P_3) + 0.09 \ln (P_2/P_3) \\ \text{se} = &\ (0.16) \quad (0.11) \qquad (0.19) \qquad (0.16) \qquad \text{RSS} = 0.364\end{aligned} \qquad (4)$$

(a) فسر المعادلتين (3) و (4) .

(b) كيف يمكنك التحقق من القيد $\alpha_1 + \alpha_2 + \alpha_3 = 1$ ؟ وضح خطواتك الحسابية اللازمة لذلك .

28.8 قدر نموذج أسعار أصول رأس المال (*CAPM*) . في الفقرة 6.1 درسنا بشكل مختصر نموذج سعر أصل رأس المال المعروف في نظرية الأوراق التجارية الحديثة . في تحليل تطبيقي ، تم تقدير CAPM على مرحلتين كالتالي :

مرحلة I (**انحدار سلاسل زمنية**) . بالنسبة لـ N قطاع موجودين في القيمة ، ثم تقدير الانحدار التالي :

$$R_{it} = \hat{\alpha}_i + \hat{\beta}_1 R_{mt} + e_{it} \qquad (1)$$

حيث: R_{it} و R_{mt} هي معدلات العائد على القطاع، وعلى سوق الأوراق التجارية (مثلاً S&P 500) في السنة t، β_i كما عرفناه سابقاً، هو Beta أو معامل تغيير السوق للقطاع i و e_{it} هو البواقي. ويوجد لدينا N من هذه الانحدارات واحدة لكل قطاع، وبالتالي لدينا N تقدير للـ β_i.

مرحلة II **(انحدار القطاع العرضي)**. في هذه المرحلة نقوم بعمل انحدار على N قطاع كالتالي:

$$\overline{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2 \hat{\beta}_i + u_i \qquad (2)$$

حيث: $\overline{R}_i$ هي المتوسط أو معدل متوسط العائد للقطاع i محسوب خلال الفترة الزمنية الموجودة في المرحلة I، $\hat{\beta}_i$ مقدرة beta من انحدار المرحلة الأولى، و u_i هو مقدار البواقي.

بمقارنة انحدار المرحلة الثانية (2) معادلة CAPM رقم (2.1.6) يمكن كتابة التالي:

$$\mathrm{ER}_i = r_f + \beta_i(\mathrm{ER}_m - r_f) \qquad (3)$$

حيث: r_f = معدل العائد الخالي من المخاطر، رأينا أن $\hat{\gamma}_1$ هو تقدير r_f، و $\hat{\gamma}_2$ هو تقدير ($\mathrm{ER}_m - r_f$) والذي يمثل خطر السوق الأولي.

وبالتالي، في الاختبار التطبيقي (CAPM)، $\overline{R}_i$ و $\hat{\beta}_i$ قيِّم استخدامهما كمقدرين لكل من ER_i و β_i على الترتيب.

بافتراض تحقق CAPM، إحصائياً دعنا نعتبر التالي:

$$\hat{\gamma}_1 = r_f$$

$$\hat{\gamma}_2 = R_{mt} - r_f$$

والتي يعتبر مقدر لـ $\mathrm{ER}_m - r_f$

دعنا الآن نعتبر النموذج البديل التالي:

$$\overline{R}_i = \hat{\gamma}_1 + \hat{\gamma}_2 \hat{\beta}_i + \hat{\gamma}_3 s_{ei}^2 + u_i \qquad (4)$$

حيث: s_{ei}^2 = تباين بواقي القطاع، من المرحلة الأولى للانحدار. وبالتالي إذا تحققت الـ CAPM، $\hat{\gamma}_3$ يجب ألا يختلف إحصائياً عن الصفر.

لاختبار CAPM، قام Levy بإجراء الانحدارين (2) و(4) على عينة من 101 مخزون خلال الفترة 1998 إلى 1968 وحصل على النتائج التالية : (*)

$$\hat{\bar{R}}_i = 0.109 + 0.037\beta_i \qquad (2)'$$
$$(0.009) \quad (0.008)$$
$$t = (12.0) \quad (5.1) \qquad R^2 = 0.21$$

$$\hat{\bar{R}}_i = 0.106 + 0.0024\hat{\beta}_i + 0.201 s^2_{\epsilon i} \qquad (4)'$$
$$(0.008) \quad (0.007) \quad (0.038)$$
$$t = (13.2) \quad (3.3) \quad (5.3) \qquad R^2 = 0.39$$

(a) هل هذه النتائج مدعمة لـ CAPM ؟

(b) هل هناك فائدة من إضافة $s^2_{\epsilon i}$ للنموذج ؟ كيف يمكنك معرفة ذلك؟

(c) إذا تحقق CAPM، $\hat{\gamma}_2$ في '(2) يمكن استخدامها كتقريب لمتوسط قيمة المعدل الخالي من المخاطر r_f. القيمة المقدرة هي 10.9%. هل هذه النسبة تبدو منطقية كمقدر لمعدل العائد الخالي من المخاطر خلال فترة الدراسة 1948 إلى 1968 ؟ (يمكنك الاستناد إلى معدل العائد على سندات الخزانة ، أو أي أصول مماثلة خالية من المخاطر).

(d) إذا تحقق CAPM، فإن دفعة مخاطرة السوق $(\bar{R}_m - r_f)$ من '(2) تساوي تقريباً 3.7% . إذا افترضنا أن r_f تساوي تقريباً 10.9%، فإن هذا يعني أن $\bar{R}_m$ من القيمة يكون مساوياً تقريباً لـ 14.6% . هل هذا التقدير يبدو منطقيًا من وجهة نظرك؟

(e) ما الذي يمكنك قوله بوجه عام عن الـ CAPM ؟

29.8 بالإشارة إلى تمرين 12.c.7 والآن بعد أن تعرفت على أدوات التحليل اللازمة، أي اختبار أو اختبارات يمكن أن تستخدم للاختيار ما بين نموذجين مختلفين؟ وضح الحسابات اللازمة لذلك. لاحظ أن المتغير التابع ليس هو نفس المتغير في النموذجين.

30.8 لمعرفة ما إذا كان هناك ثابت على العائد في الاقتصاد المكسيكي خلال فترة الدراسة أم لا.

(*) H. Levy, "Equilibrium in an Imperfect Market: A constraint on the number of securities in the portfolio," American Economic Review, vol. 68, no. 4, September 1978, pp. 643–658.

31.8 بالعودة إلى مثال وفيات الأطفال الذي ناقشناه عدة مرات سابقًا. في الانحدار (2.6.7) نقوم بعمل انحدار لوفيات الأطفال (CM) على GNP للفرد (PGNP) ومعدل معرفة القراءة والكتابة في النساء (FLR). الآن دعنا نضيف متغيراً جديداً للنموذج، وهو معدل الخصوبة الإجمالي (TFR). البيانات الخاصة بهذه المتغيرات معطاة في الجدول (4.6). بإعادة عمل الانحدار (2.6.7) بعد إضافة المتغير الجديد حصلنا على البيانات التالية:

$$1.\ \widehat{CM}_i = 263.6416 - 0.0056\ \text{PGNP}_i - 2.2316\ \text{FLR}_i \qquad (2.6.7)$$
$$\text{se} = (11.5932) \quad (0.0019) \qquad (0.2099) \qquad R^2 = 0.7077$$

$$2.\ \widehat{CM}_i = 168.3067 - 0.0055\ \text{PGNP}_i - 1.7680\ \text{FLR}_i + 12.8686\text{TFR}_i$$
$$\text{se} = (32.8916) \quad (0.0018) \qquad (0.2480) \qquad (?)$$
$$R^2 = 0.7474$$

(a) كيف يمكنك تفسير معامل TFR ؟ مبدئياً، هل تتوقع علاقة موجبة أم سالبة بين كل من CM و TFR ؟ علل إجابتك.

(b) هل معاملٍ كل من PGNP و FR تغييرات بين المعادلتين السابقتين؟ وإذا كان ذلك صحيحًا، ما هو السبب (أو الأسباب) التي أدت إلى مثل هذا التغيير؟ هل هذا الفرق معنوي إحصائياً؟ ما هو الاختبار الذي ستستخدمه ولماذا؟

(c) كيف يمكنك الاختيار بين نموذج (1) أو نموذج (2)؟ ما هو الاختبار الإضافي الذي ستستخدمه للإجابة على هذا السؤال؟ وضح الحسابات اللازمة لذلك.

(d) الخطأ المعياري لمعامل TFR ليس معطى. هل يمكنك حسابه؟ (ملاحظة: استخدم العلاقة بين توزيع t وتوزيع F).

32.8 بالإشارة إلى تمرين 7.1، والذي يعطي بيانات عن أثر الدعاية والإعلان والتكلفة الخاصة بهما لعينة من 21 مشروعًا في تمرين 11.5 سُئلت من قبل عن عمل رسم بياني لتوضيح هذه البيانات واستعراضها لتحديد النموذج المناسب لها بناء على العلاقة بين إثر الإعلان وتكلفته. دع Y تمثل أثر الدعاية والانطباع الذى تتركه عند المتلقى، و X تكلفة الدعاية، الانحدار الخاص بذلك تم الحصول عليه كالتالي:

نموذج I :

$$\hat{Y}_i = 22.163 + 0.3631X_i$$
$$se = (7.089) \quad (0.0971) \qquad r^2 = 0.424$$

نموذج II :

$$\hat{Y}_i = 7.059 + 1.0847X_i - 0.0040X_i^2$$
$$se = (9.986) \quad (0.3699) \quad (0.0019) \qquad R^2 = 0.53$$

(a) فسر النموذجين.

(b) أي منهما أفضل؟ لماذا؟

(c) ما هو الاختبار (أو الاختبارات) الإحصائية للاختيار بين النموذجين؟

(d) هل هناك «عوائد متناقصة» لتكلفة الإعلان (أي أنه بعد مستوى معين من تكلفة الإعلان [وتسمى مستوى التشبع] لا يوجد مبرر من الاستمرار في تحمل تلك التكلفة)؟ هل يمكن تحديد هذا المستوى؟ وضح الخطوات الحسابية اللازمة لذلك.

33.8 في انحدار (4.9.7)، وضحنا النتائج التي حصلنا عليه من دالة إنتاج Cobb - Douglas والتي تم توفيقها لقطاع الزراعة التايواني خلال الفترة 1958 إلى 1972. بناء على نتائج هذا الانحدار، وضح ما إذا كان هناك ثابت على العائد في هذا القطاع أم لا، **يمكنك القيام بذلك باستخدام التالي:**

(a) اختبار t المعطي في (4.7.8)، مع ملاحظة أن التغاير بين المقدرين الخاصين بالميل هو 0.03843-.

(b) اختبار F المعطى في (9.7.8).

(c) هل هناك فرق بين نتيجة كل من الاختبارين السابقين؟ وما هو الاستنتاج الذي حصلت عليه بخصوص قطاع الزراعة في تايوان خلال فترة الدراسة؟

34.8 بالرجوع إلى انحدار الادخار – الدخل المعطى في الفقرة 8.8 افترض أننا قسمنا العينة على فترتين: 1970 إلى 1982 و 1983 إلى 1995. باستخدام اختبار Chow، حدد ما إذا كان هناك تغيير هيكلي في انحدار الادخار – الدخل خلال هاتين العينتين السابقتين أم لا.. قارن نتائجك مع النتائج المعطاة في الفقرة 8.8، ما هو الاستنتاج العام الذي يمكنك الوصول إليه والخاص بمدى حساسية اختبار chow لاختبار نقطة التغيير أو التحول التي نفصل بها العينة إلى عينتين (أو أكثر)؟

ملحق A8(*) APPENDIX

اختبار نسبة الإمكان (LR): LIKELIHOOD RATIO (LR) TEST

اختبار LR يعتمد على فكرة الإمكان الأعظم (ML) التي سبق وناقشناها في ملحق A4، والتي أوضحنا فيها كيف يمكن الحصول على مقدرات ML لنموذج انحدار ثنائي المتغيرات. الطريقة يمكن تعميمها بشكل مباشر لنموذج الانحدار متعدد المتغيرات. تحت صحة الفرض الخاص بتبعية حد الخطأ u_i للتوزيع الطبيعي، أثبتنا أنه بالنسبة لنموذج الانحدار ثنائي المتغيرات، مقدرات OLS و ML لمعاملات الانحدار هي مقدرات متساوية، وإن كانت مقدرات تباين الأخطاء مختلفة. في حين أن مقدر OLS لـ σ^2 هو $\sum\hat{u}_i^2/(n-2)$ مقدر ML هو $\sum\hat{u}_i^2/n$، الأول مقدر غير متحيز، أما الأخير فهو مقدر متحيز، مع ملاحظة أنه في العينات الكبيرة يتلاشى هذا التحيز. هذا هو نفس الحال في الانحدار المتعدد. **لتوضيح ذلك دعنا نفترض النموذج ثلاثي المتغيرات التالي:**

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (1)$$

وفقاً للمعادلة (5) في ملحق A4، دالة لوغاريتم الإمكان لنموذج (1) يمكن كتابتها كالتالي:

$$\ln \text{LF} = -\frac{n}{2}\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum(Y_i - \beta_1 - \beta_2 X_{2i} - \beta_3 X_{3i})^2 \qquad (2)$$

وكما هو موضح في ملحق A4، بأخذ تفاضل هذه الدالة بالنسبة لـ β_1، β_2، β_3، σ^2 ومساواة الناتج بالصفر، وحل هذه المعادلات، نحصل على مقدرات ML لهذه القيم. مقدرات ML لـ β_1، β_2 و β_3 ستكون متساوية تماماً مع مقدرات OLS والمعطى في المعادلات (6.4.7) إلى (8.4.7)، لكن تباين الخطأ سيكون مختلفًا، حيث إن مجموع مربعات البواقي سيكون مقسومًا على n بدلاً من $n-3$ في حالة الـ OLS.

والآن دعنا نفترض أن فرضنا العدمي H_0 هو أن معامل X_3، β_3، يساوي الصفر. في هذه الحالة، فإن لوغاريتم LF المعطى في (2) سيصبح كالتالي:

$$\ln \text{LF} = -\frac{n}{2}\sigma^2 - \frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum(Y_i - \beta_1 - \beta_2 X_{2i})^2 \qquad (3)$$

(*) اختياري.

المعادلة (3) معروفة باسم دالة لوغاريتم الإمكان المقيدة (RLLE)، حيث إنها مقدرة وفقاً لقيد مسبق، وهو B3 يساوي الصفر، وبالتالي فإن المعادلة (1) معروفة باسم لوغاريتم LF غير المقيد (ULLF)، حيث إنه لا توجد أي قيود مسبقة على المعامل. لاختبار صحة القيد المسبق على β_3 والذي يجعلها مساوية للصفر، اختبار LR يتم وفقاً لإحصاء الاختبار التالي :

$$\lambda = 2\ (\text{ULLF} - \text{RLLF}) \qquad (4)^*$$

حيث ULLF و RLLF هما، على الترتيب، دالة لوغاريتم الإمكان غير المقيدة [المعادلة (1)] ودالة الإمكان المقيدة [المعادلة (3)]. إذا كان حجم العينة كبيراً، يمكن إثبات أن إحصاء الاختبار λ والمعطى في (4) يتبع توزيع كاي - التربيعي (X^2) بدرجات حرية تساوي عدد القيود المفروضة في الفرض العدمي، وتساوي 1 في مثالنا الحالي.

الفكرة الرئيسية وراء اختبار LR هي فكرة بسيطة يمكن التعبير عنها كالتالي : إذا كانت القيود المسبقة صحيحة، فإن لوغاريتم (LF) المقيد وغير المقيد يجب ألا يختلفان. وفي هذه الحالة، فإن (λ) في (4) سيساوي الصفر. ولكن إذا كان ذلك غير حادث، فإن قيمتي الـ F، ستكونان مختلفتين وحيث إننا في العينات كبيرة الحجم، نعرف أن λ تتبع توزيع كاي - التربيعي، فإنه يمكننا معرفة المعنوية الإحصائية للفرق بين القيمتين، عند مستوى معنوية مثلاً 1 أو 5% ولأي مستوى معنوية أخرى، يمكننا الحصول على p-value لقيمة λ المقدرة.

دعنا الآن نستعرض اختبار LR على مثال وفيات الأطفال، إذا قمنا بعمل انحدار لوفيات الأطفال (CM) على GNP للفرد (PGNP) ومعدلات القراءة والكتابة عند النساء (FLR) كما فعلنا من قبل في (1.2.8)، سنحصل على ULLF مساويًا لـ 328.1012– ولكن إذا قمنا بعمل انحدار لـ CM على PGNP فقط سنحصل على RLLF مساويًا لـ 361.6396–. إذا استخدمنا القيم المطلقة (أي استبعدنا الإشارة)، فإن العينة الأولى أصغر من القيمة الأخيرة، والذي يبدو منطقيًا، حيث إن لدينا متغيراً إضافيًا في الحالة الأولى.

(*) هذا التعبير يمكن كتابته على الشكل $-2(\text{RLLF}-\text{VLLF})$ أو $2 \ln(\text{RLF}/\text{ULF})$.

السؤال الآن عن مدى أهمية إضافة المتغير FLR. إذا كان غير مهم، فإن LLF المقيدة وغير المقيدة، يجب ألا يختلفان كثيراً، ولكن إذا كان مؤثراً فمن المفروض أن يختلفا. لدراسة ما إذا كان هذا الفرق له معنوية إحصائية أم لا، نستخدم الآن اختبار LR المعطى في (4) كالتالي:

$$\lambda = 2[-328.1012 - (-361.6396)] = 67.0768$$

هذه القيمة تقاربياً لها توزيع كاي - التربيعي. بدرجة حرية 1 (حيث إن لدينا قيداً واحداً فقط خاصاً بحذف المتغير FLR من النموذج الكلي). قيمة p-value التي نحصل عليها من كاي - التربيعي لدرجة حرية 1 تساوي الصفر تقريباً، مما يعني أن المتغير FLR لا يجب حذفه من النموذج. بمعنى آخر، الانحدار المقيد في المثال الحالي غير صحيح.

نظراً للصعوبة الرياضية لاختباري LM وWald، لن نستعرضهما في المناقشة الحالية. لكن كما لاحظنا من قبل، فإن اختبارات LR و Wald و LM تعطي تقاربياً نفس النتائج، والاختيار ما بين هذه الاختبارات يعود إلى الإمكانية الحسابية المتوافرة لذلك.

الجزء الثاني

تحرير (تخفيف) فروض النموذج التقليدي

RELAXING THE ASSUMPTIONS OF THE CLASSICAL MODEL

في الجزء I ، استعرضنا نموذج الانحدار الخطي الطبيعي التقليدي ، وأوضحنا كيف يمكن من خلاله التعامل مع المشكلتين المزدوجتين للاستدلال الإحصائي ، وهما التقدير ، واختبارات الفروض بالإضافة إلى مشكلة التنبؤ.

ولكن تذكر أن هذا النموذج قائم على مجموعة مبسطة من الفروض كالتالي :

الفرض 1 . نموذج انحدار خطي في المعلمات .

الفرض 2 . قيم المتغيرات المنحدر عليه ، X's، ثابتة في العينات المتكررة .

الفرض 3 . بمعلومية الـ X's، القيمة المتوسطة للخطأ u_i تساوي الصفر .

الفرض 4 . بمعلومية الـ X's، تباين u_i ثابت .

الفرض 5 . بمعلومية الـ X's، لا يوجد ارتباط ذاتي للأخطاء .

الفرض 6 . إذا كانت X's عشوائية ، فإن مقدار الخطأ و المتغيرات X's العشوائية مستقلة أو على الأقل غير مرتبطة .

الفرض 7 . عدد المشاهدات لابد أن يكون أكبر من عدد المتغيرات المنحدر عليها .

الفرض 8 . لابد من وجود تباين كاف في القيم الخاصة بالمتغيرات المنحدر عليها .

الفرض 9 . نموذج الانحدار محدد بشكل سليم .

الفرض 10 . لا توجد علاقة خطية تامة بين المتغيرات المنحدر عليها (أي لا يوجد ارتباط خطي متعدد) .

الفرض 11 . مقدار الخطأ العشوائي u_i يتبع التوزيع الطبيعي .

وقبل الدخول في مزيد من التفاصيل ، دعنا نلاحظ أن معظم الكتب تتناول عدداً من الفروض أقل من 11 فرضاً . فعلى سبيل المثال ، الفرضان 7 و 8 قيم افتراضهما ضمنياً ولا ينص عليهما صراحةً . وقد قررنا كتابة هذين الفرضين بشكل صريح كضرورة للتفرقة بين الفروض اللازمة لجعل الـ OLS لها خصائص إحصائية مميزة (مثل BLUE) ، والفروض اللازمة لجعل الـ OLS مقيدة لاستخدامهما . فمثلاً ، مقدرات OLS تعتبر BLUE حتى إذا لم يتحقق الفرض 8 . لكن في هذه الحالة ، فإن الخطأ المعياري لمقدرات الـ OLS سيكون كبيراً مقارنةً بقيمتها (مما يعني أن نسبة t ستكون صغيرة) ، مما يجعل هناك صعوبة في تحديد أهمية متغير أو أكثر من المتغيرات المنحدر عليها في تفسير مجموع المربعات .

وكما سبق أن أشرنا ، هناك مشكلتان رئيستان تظهران عند تطبيق نموذج الانحدار الخطي التقليدي وهما : (1) مشاكل ترجع إلى فروض توصيف النموذج والأخطاء u_i و (2) مشاكل ترجع إلى فروض خاصة بالبيانات[1]. الفروض 1 ، 2 ، 3 ، 4 ، 5 ، 9 ، 12 تقع في إطار المجموعة الأولى ، أما الفروض 6 ، 7 ، 8 و 10 فهي تقع في إطار المجموعة الثانية . بالإضافة إلى أن مشاكل البيانات الخاصة بالقيم الشاذة (مشاهدات غير معتادة) ، وأخطاء القياس في البيانات تقع أيضاً في إطار المجموعة الثانية .

بالنسبة للمشاكل الناشئة عن فروض خاصة بالأخطاء وتوصيف النموذج ، يوجد لدينا ثلاثة أسئلة رئيسة وهي :

---

(1) G. Barrie Wetherill, *Regression Analysis with Applicatins*, Chapman and Hall, New York, 1986, 99. 14–15.

1 – متى يمكن القول بأننا بعدنا بشكل كبير عن فرض معين ، مما يؤدي لحدوث مشكلة؟ فمثلاً ، إذا كانت ui معيناً لا تتبع بالضبط التوزيع الطبيعي ، ما هو المستوى المقبول من البعد عن التوزيع الطبيعي والذي بعده تفقد مقدرات الـ OLS صفة الـ BIUE؟

2 – كيف يمكننا من الناحية الفعلية التأكد من أي فرض من الفروض السابقة ، وما إذا كان فعلياً متحققاً أم غير متحقق؟ بمعنى كيف مثلاً يمكن التحقق من أن الأخطاء تتبع التوزيع الطبيعي أم لا عند التحقيق على حالة معينة؟

لقد سبق وناقشنا اختبار التوزيع الطبيعي لـ Anderson - darling ، وكاي – التربيعي و Jarque - Bera .

3 – ما هي المقاييس الإصلاحية التي يمكن أن نستخدمها إذا كان واحداً أو أكثر من الفروض غير متحقق؟ فمثلاً ، إذا كان فرض ثبات التباين غير متحقق في أحد التطبيقات ، ماذا نفعل في هذه الحالة؟

أما بالنسبة للمشاكل المتعلقة بفروض البيانات ، فإن لدينا أسئلة مناظرة كالتالي :

1 – كيف نعرف حدة المشكلة التي نواجهها؟ فمثلاً هل الارتباط المتعدد له خطورة شديدة لدرجة أنه يعيق عملية التقدير والاستدلال؟

2 – كيف يمكننا أن نتحقق من خطورة مشكلة البيانات؟ فمثلاً كيف يمكن أن نحدد كيفية التعامل مع بعض المفردات التي تمثل قيماً شاذة داخل البيانات ، وتؤدي إلى صعوبة في التحليل ، هل نتعامل معها بالحذف أم بالاحتفاظ؟

3 – هل بعض مشاكل البيانات من الممكن علاجها؟ فمثلاً ، هل ممكن للباحث أن يحصل على البيانات الأصلية بمعرفة مصدر أخطاء القياس فيها؟ للأسف لا توجد إجابات وافية لكل هذه الأسئلة . وما ستقوم به في الجزء الثاني من هذا الكتاب هو دراسة بعض الفروض السابق ذكرها بمزيد من التفصيل ، وسنستعرض البعض وليس الكل . وبالتحديد فلن نناقش بالتفصيل التالي :

الفروض 2، 3، 6 و 11. الفروض 2 و 6 والخاصة بالمتغيرات المنحدرة عليه سواء ثابتة العشوائية . تذكر أن تحليل الانحدار مبني على أساس أن المتغيرات

المنحدر عليه غير عشوائية ويفترض ثباتها في العينات المتكررة. وهناك سبب وجيه لذلك. فعلى عكس العلوم الطبيعية – وكما لاحظنا في الفصل الأول – الاقتصاديون بوجه عام لا يستطيعون التحكم في البيانات التي يستخدمونها. ففي كثير من الأحيان، يعتمد الاقتصاديون على بيانات ثانوية، بمعنى أنها بيانات تم تجميعها بواسطة جهة أخرى مثل الحكومة أو المؤسسات الخاصة. وبالتالي فأسلوب التعامل العملي مع هذه البيانات، هو افتراض ثبات قيم المتغيرات المفسرة، فحتى وإن كانت المتغيرات نفسها ممكن أن تكون عشوائية أو متغيرة. وبالتالي، فإن نتائج تحليل الانحدار مشروطة بقيم هذه المتغيرات.

ولكن افترض الآن أنه لا يمكن التعامل مع $X's$ على أنها قيم ثابتة غير عشوائية. في مثل هذه الحالة، لابد من التعامل مع متغيرات منحدر عليها متغيرة أو عشوائية. ويصبح الموقف أكثر تعقيداً. فالـ u_i وفقاً للفروض السابقة تعتبر متغيرات عشوائية. وإذا كانت $X's$ عشوائية أيضاً، لابد من معرفة كيفية توزيع $X's$ و u_i. إذا رغبنا في تحقيق الفرض 6 (أي أن $X's$، على الرغم من أنها عشوائية، موزعة بشكل مستقل عن الـ u_i أو على الأقل غير مرتبطة مع u_i)، لا نستطيع بشكل عملي الاستمرار مع فكرة التعامل مع $X's$ كمتغيرات غير عشوائية. وكما ذكر Kmenta فإن:

تحرير الفرض القائل بأن X متغير غير عشوائي واستبداله بالفرض القائل بأن X متغير عشوائي ولكنه مستقل عن الـ u لا يؤثر على الخصائص المرغوب فيها لتقديرات المربعات الصغرى[2].

وبالتالي سنحتفظ بالفرض 2 أو الفرض 6 حتى نأتي إلى نماذج المعادلات الآنية في الجزء VI [3].

---

(2) Jan Kmenta, Elements of Econometrics, 2d ed., Macmillan, New York, 1986, p. 338. (Emphasis in the original).

(3) هناك ملاحظة فنية هنا يجب الإشارة إليها وهي أنه بدلاً من الفرض أقوى والخاص باستقلال $X's$ و u من الممكن استخدام فرض أضعف وهو عدم ارتباط قيم متغيرات الـ X و u لحظياً (أي عند نفس النقطة الزمنية). في هذه الحال ممكن أن تكون مقدرات OLS متحيزة ولكنها منسقة. أي أنه مع زيادة حجم العينة يقترب المقدر من القيمة الحقيقية. وبالتالي إذا كانت $X's$ و u مرتبطين لحظياً فإن متغيرات OLS متغيرة وغير منسقة. في الفصل 17 نستعرض المتغيرات الوسيطة وكيفية استخدامها للحصول على مقدرات منسقة في مثل هذه الحالة.

الفرض 3. توقع u_i يساوي الصفر. دعنا نتذكر نموذج الانحدار الخطي، والذي يشتمل على k متغير كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \qquad (1)$$

دعنا الآن نفترض التالي:

$$E(u_i \mid X_{2i}, X_{3i}, \ldots, X_{ki}) = w \qquad (2)$$

حيث w ثابت، في النموذج القياسي $w = 0$، ولكن دعنا الآن نفترض أنه فقط عبارة عن ثابت.

بإدخال التوقع الشرطي على (1)، نحصل على

$$\begin{aligned} E(Y_i|X_{2i}, X_{3i}, \ldots, X_{ki}) &= \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + w \\ &= (\beta_1 + w) + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \quad (3) \\ &= \alpha + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} \end{aligned}$$

حيث: $\beta_1 + w = \alpha$ ويمكن ملاحظة أنه عند إدخال التوقع على المعادلة السابقة، فإن $X's$ يتم التعامل معها على أنها ثابتة.

وبالتالي إذا كان الفرض 3 غير متحقق، فإننا لا نستطيع تقدير الجزء المقطوع من المحور الصادي β_1، فالذي حصلنا عليه α يمثل β_1 و $E(u_i) = w$. وبالتالي للاختصار، يمكن القول إننا حصلنا على مقدر متغير لـ β_1.

ولكن كما سبق ولاحظنا في مواطن عديدة من قبل، في العديد من التطبيقات، فإن الجزء المقطوع من المحور الصادي β_1 له أهمية قليلة. فالقيم التي لها أهمية أكبر هي معاملات الميل، والتي لن تتأثر حتى إذا لم يتحقق الشرط 3. [4] بالإضافة إلى ذلك، فإنه في العديد من التطبيقات لا يوجد معنى حقيقي للجزء المقطوع من المحور الصادي.

---

(4) من المهم ملاحظة أن هذه العبارة سليمة فقط في حالة أن $E(u_i) = w$ لكل i. وبالتالي إذا كان $E(u_i) = w_i$ ولكنه يختلف من i لأخرى، فإن معاملات الميل الجزيئية قد تكون متحيزة وغير متسقة وفي هذه الحالة عدم تحقق الفرض 3 يصبح له أهمية كبيرة. للاثبات النظري ومزيد من التفاصيل، انظر Peter Schmidt, Econometrics, Marcel Dekker, New York, 1976, pp. 36–39.

فرض 11: اعتيادية u. هذا الفرض غير مهم إذا كنا مهتمين فقط بالتقدير، كما لاحظنا في الفصل (3)، مقدرات OLS تعتبر BIUE بغض النظر عن تباعية u_i للتوزيع الطبيعي من عدمها.

وعلى الرغم من ذلك، فإنه مع تحقق فرض اعتيادية الأخطاء، نستطيع القول بأن مقدرات OLS لمعاملات الانحدار تتبع التوزيع الطبيعي.

أي أن $(n-k)\frac{\hat{\sigma}^2}{\sigma^2}$ له توزيع كاي-التربيعي، وبالتالي يمكننا استخدام اختبارات t، F والخاصة بالعديد من الفروض الإحصائية وبدون الاهتمام بحجم العينة.

ولكن ماذا سيحدث إذا كانت u_i لا تتبع التوزيع الطبيعي؟ لابد إذن من الاعتماد على نظرية النزعة المركزية، تذكر أن نظرية السرعة المركزية هي أيضاً التي استخدمت من قبل لتفسير استخدام فرض الاعتيادية منذ البداية كالتالي:

اذا كانت الأخطاء $[u_i]$ مستقلة وموزعة بشكل متماثل بتوقع يساوي الصفر، وتباين [ثابت] σ^2 وإذا كانت المتغيرات المفسرة ثابتة في العينات المتكررة، فإن مقدرات OLS للمعاملات تتبع تقاربياً التوزيع الطبيعي بتوقع يساوي β's المناظرة لها.[5]

وبالتالي خطوات اختبارات F و t التقليدية سليمة تقاربياً، أي في أحجام العينات الكبيرة، ولكن ذلك غير متحقق في العينات المحدودة أو الصغيرة.

وبالتالي في الواقع إذا كانت الأخطاء لا تتبع التوزيع الطبيعي، فإن تباعية مقدرات الـ OLS بشكل تقاربي للتوزيع الطبيعي (مع ثبات التباين وافتراض ثبات X's) يصعب عملياً من الناحية الاقتصادية، حيث إن توافر عينات كبيرة الحجم أمر يصعب حدوثه في الواقع.

ولهذا، فإن فرض الاعتيادية يصبح له أهمية كبيرة في اختبارات الفروض والتنبؤ.

---

(5) Henri Theil, Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, N.J., 1978, p. 240. It must be noted the assuptions of fixed X's and constant σ^2 art crucial for this result.

وبالتالي ، إذا وضعنا في الاعتبار المشكلة المزدوجة للتقدير واختبارات الفروض في الاعتبار، مع اعتبار الحقيقة الخاصة بأن أحجام العينات الصغيرة هي القاعدة وليست الاستثناء في معظم التحاليل الاقتصادية ، فإنه يجب الاستقرار في الاعتماد على فرض الاعتيادية .[6]

وبالتالي ، فإن ذلك يعني بالطبع أنه في حالة العينات المحدودة ، لابد من أن نختبر صراحة فرض الاعتيادية . قد سبق واستعرضنا اختبارات Anderson-Darling و Jarque-Bera للاعتيادية .

وننصح الباحث بشدة أن يطبق هذه الاختبارات أو أي اختبارات أخرى لفرض الاعتيادية على بواقي الانحدار. مع الوضع في الاعتبار بأنه في العينات المحدودة الإحصاء المعتاد لـ t و F لن يتبع توزيع t و F إذا كان فرض الاعتيادية غير متحقق .

يبقى لنا الآن الفروض 1، 4، 5، 7، 8، 9، 10 . الفروض 7، 8، 10 مرتبطة ببعضها البعض ، وتمت مناقشتها في الفصل الخاص بالارتباط المتعدد (الفصل 10) . الفرض 4 تمت مناقشته في الفصل الخاص بالارتباط الذاتي (الفصل 12) . الفرض 9 يناقش في الفصل الخاص بتوصيف النموذج والاختبارات (الفصل 13) . ونظراً لطبيعة الفرض 1 الخاصة والمتطلبات الرياضية التي تلزم مناقشته ، سيتم استعراضه كموضوع خاص في الجزء III (الفصل 14) لأسباب تنظيمية ، في كل من هذه الفصول ، سنتبع نمطًا معينًا لاستعراض المواضيع وهنا النمط يكون كالتالي :

1 – تحديد طبيعة المشكلة .

2 – اختبار عواقبها .

3 – اقتراح طرق للتعرف عليها .

---

(6) ملاحظة عابرة ، أثر البعد عن فرض الاعتيادية والمواضيع المتعلقة بذلك يتم مناقشتها في إطار Robust estimatim وهو موضوع خارج إطار هذا الكتاب .

4 - اعتبار بعض المقاييس الأخرى والتي قد تؤدي إلى مقدرات تحقق الخصائص الإحصائية المطلوبة والتي ناقشناها في الجزء I.

ويجب ملاحظة هنا - كما سبق وذكرنا من قبل - فإنه لا توجد إجابات وافية لكل المشاكل السابق ذكرها والمخالفة لفروض الـ CLRM. بالإضافة إلى أنه أحياناً يكون هناك أكثر من حل لمشكلة معينة، ويكون من غير الواضح أي من هذه الحلول هو الأفضل. هذا بالإضافة إلى أنه في بعض التطبيقات، قد يكون هناك عدم تحقق لأكثر من فرض واحد من فروض الـ CLRM.

فمثلاً تحيز التوصيف، الارتباط المتعدد، وعدم ثبات التباين، قد توجد جميعاً في تطبيق عملي واحد.

ولا يوجد اختبار واحد قاطع قادر على حل كل هذه المشاكل آنياً. (7)

والأكثر من ذلك، فقد يكون هناك اختبار معين مناسب في مرحلة زمنية معينة، ولا يصلح للاستخدام في وقت لاحق، فقد يكون باحث آخر قد أظهر الجوانب السلبية لهذا الاختبار مع مرور الزمن.

ولكن هكذا يتقدم العلم، والاقتصاد القياسي غير مستثنى من ذلك.

---

(7) هذا ليس ناتجًا من عدم المحاولة. انظر A. K. Bera and C. M. Jarque, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence," Economic Letters, vol. 7, 1981, pp. 313–318.

الفصل التاسع

نماذج الانحدار ذات المتغير الوهمي
DUMMY VARIABLE REGRESSION MODELS

في الفصل (1) ، ناقشنا باختصار، أنواع المتغيرات الأربعة التي يواجهها الباحث كثيراً في التحليل التطبيقي. . وهي كالتالي: مقياس النسبة، مقياس الفترة، المقياس الترتيبي، والمقياس الاسمى. أنواع المتغيرات التي تعلمناها في الفصول السابقة كانت في الأغلب مقاييس النسبة، ولكن هذا يجب ألا يعطي انطباعًا بأن نماذج الانحدار لا تتكامل إلا مع المتغيرات ذات مقياس النسبة. فنماذج الانحدار يمكن تطبيقها أيضاً مع الأنواع الأخرى السابق ذكرها.

في هذا الفصل، سنعتبر نماذج انحدار لا تحتوي فقط على متغيرات ذات مقياس النسبة، ولكن أيضاً مع المتغيرات ذات المقياس الاسمى. مثل هذه المتغيرات معروفة بأسماء عديدة منها، متغيرات المؤشر، المتغيرات الطبقية، المتغيرات النوعية، أو المتغيرات الوهمية[1].

1.9 طبيعة المتغيرات الوهمية :
THE NATURE OF DUMMY VARIABLES

في نموذج الانحدار المتغير التابع، أو المنحدر يتأثر كثيراً ليس فقط بمتغيرات ذات مقياس النسبة (مثل الدخل، الناتج، الأسعار، التكلفة، الطول، درجة الحرارة)، ولكن أيضاً بمتغيرات بطبيعتها نوعية، أولها مقياس اسمى مثل النوع، العرق، اللون، الديانة، الجنسية، المنطقة الجغرافية، المعتقدات السياسية و الاتجاهات

(1) سنناقش متغيرات المقياس الترتيبي في الفصل (15) .

الحزبية. فعلى سبيل المثال، بافتراض ثبات كل العوامل الأخرى، فإن السيدات العاملات وجد أن دخلهن أقل من نظرائهن من ذوات البشرة البيضاء[2].

هذا قد يكون له علاقة بالتمييز في النوع أو العرق، ولكن بغض النظر عن الأسباب، فإن المتغيرات النوعية مثل النوع والعرق يبدو أن لها تأثيرًا على المتغير المنحدر، ويبدو أهمية إدخالها إلى النموذج مع باقي المتغيرات المفسرة أو المنحدر عليها.

وبما أن هذه المتغيرات تمثل مؤشرًا لظهور أو غياب صفة ما مثل ذكر أو أنثى، أبيض أو أسود، كاثوليك أو غير الكاثوليك، حزب ديمقراطي أو جمهوري، هؤلاء جميعاً متغيرات ذات مقياس اسمي.

إحدى طرق التعبير عن هذه المتغيرات تكون من خلال متغيرات صناعية، والتي تأخذ القيم 1 و 0، 1 معبرة عن وجود (أو ظهور) الصفة، و0 تعني عدم وجودها. فعلى سبيل المثال 1 قد يكون مؤشرًا لأن الشخص امرأة، و 0 قد تعني رجلاً أو 1 قد تعني أن الشخص خريج جامعي، و 0 تعني أنه ليس خريج جامعي وهكذا. المتغيرات التي لها قيمتان فقط 0 ، 1 تسمى متغيرات وهمية[3]. هذه المتغيرات موجودة كضرورة لتقسيم البيانات إلى طبقات متنامية كلياً مثل ذكور وإناث.

المتغيرات الوهمية يمكن إدخالها في نموذج الانحدار مثلها مثل المتغيرات الكمية. في حقيقة الأمر، فنموذج الانحدار قد يحتوي على متغيرات منحدرة عليها تكون جميعها نوعية أو وهمية بطبيعتها. هذه النماذج يطلق عليها **نماذج تحليل التباين (ANOVA) models**[4].

---

(2) لمراجعة ما يثبت هذه العبارات انظر Bruce E. Kaufman and Julie L. Hotchkiss, The Economics of Labor Market, 5th ed., Dryden Press, New York, 2000.

(3) ليس بالضرورة أن تأخذ المتغيرات الوهمية 0 و 1 . الثنائي (0 ، 1) يمكن تحويله إلى أي ثنائي آخر بدالة خطية مثل $Z = a + bD \ (b \neq 0)$ حيث a ، b ثوابت و 0 و $D = 1$. عندما تساوي $D = 1$ فإن لدينا $Z = a + b$ وعندما تساوي $D = 0$ إن لدينا $Z = a$. وبالتالي أصبح الإحداثي (0و1) هو $(a, a + b)$ ، فمثلاً إذا كانت $a = 1$ ، $b = 2$ المتغير الوهمي سيكون $(1, 3)$. وهذا يظهر أن المتغيرات الوهمية أو النوعية ليس لها مقياس موحد للقياس . ولهذا تسمى متغيرات ذات مقياس نوعي .

(4) نماذج ANOVA تستخدم لدراسة المعنوية الإحصائية للعلاقة بين متغير منحدر كمي ومجموعة من المتغيرات الوهمية أو النوعية . وتستخدم عادة للمقارنة بين الفرق بين القيم المتوسطة لمجموعتين أو أكثر ، وبالتالي هو أكثر عمومية من اختبار الـ t الذي يستخدم للمقارنة بين متوسطات مجموعتين أو طبقتين فقط .

2.9 نماذج ANOVA : ANOVA MODELS

لشرح نماذج ANOVA دعنا نعتبر المثال التالي :

مثال 1.9

أجور المدرسين العاملين في المدارس الحكومية وفقاً للمنطقة الجغرافية :

PUBLIC SCHOOL TEACHERS' SALARIES BY GEOGRAPHICAL REGION

جدول (1.9) يعطي البيانات الخاصة بمتوسط الراتب (بالدولار) للمدرسين العاملين بمدرسة حكومية في 50 ولاية ومنطقة كولومبيا في عام 1985 . هذه المناطق الإحدى والخمسون مقسمة إلى ثلاث مناطق جغرافية : (1) الشمال الشرقي وشمال الوسط (21 ولاية إجمالياً) (2) الجنوب (17 ولاية إجمالياً) و (3) الغرب (13 إجمالياً). في الوقت الحالي يجب ألا يزعجك نمط الجدول والبيانات الأخرى المعطاة فيه .

افترض أنك تريد معرفة ما إذا كان متوسط الأجر السنوي (AAS) للمدرسين بالمدارس الحكومية يختلف وفقاً للمناطق الجغرافية الثلاث الموجودة . إذا حسبت المتوسط الرياضي البسيط لمتوسطات الأجور الخاصة بالمدرسين في الثلاث مناطق، ستجد أن هذه المتوسطات لهذه المناطق الثلاث هي كالتالي :

$24.424.14 (الشمال الشرقي وشمال الوسط)، $22.894 (الجنوب) و $26.158.62 (الغرب). هذه الأرقام تبدو مختلفة، ولكن هل هناك اختلاف معنوي بينها؟ هناك أساليب إحصائية عديدة لمقارنة قيم متوسطين أو أكثر، والتي يطلق عليها اسم تحليل التباين[5] . ولكن نفس الفرض ممكن تحقيقه في إطار تحليل الانحدار . للتعرف على ذلك، دعنا نعتبر النموذج التالي :

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_{3i} D_{3i} + u_i \qquad (1.2.9)$$

حيث : Y_i = (متوسط) الأجر لمدرس في مدرسة حكومية في الولاية i

D_{2i} = 1 إذا كانت الولاية في الشمال الشرقي أو شمال الوسط

= 0 بخلاف ذلك (أي في أي منطقة أخرى)

D_{3i} = 1 إذا كانت الولاية في الجنوب

= 0 بخلاف ذلك (أي موجودة في أي منطقة أخرى)

---

(5) للتعرف على تطبيق عملي ، انظر John Fox, Applied Regression Analysis, Linear Models, and Related Methods, Sage Publications, 1997, Chap. 8.

جدول (1.9) متوسط الراتب للمدرسين العاملين في المدارس الحكومية وفقاً للولاية ، عام 1988

AVERAGE SALARY OF PUBLIC SCHOOL TEACHERS, BY STATE, 1988

Salary	Spending	D_2	D_3	Salary	Spending	D_2	D_3
19,583	3346	1	0	22,795	3366	0	1
20,263	3114	1	0	21,570	2920	0	1
20,325	3554	1	0	22,080	2980	0	1
26,800	4642	1	0	22,250	3731	0	1
29,470	4669	1	0	20,940	2853	0	1
26,610	4888	1	0	21,800	2533	0	1
30,678	5710	1	0	22,934	2729	0	1
27,170	5536	1	0	18,443	2305	0	1
25,853	4168	1	0	19,538	2642	0	1
24,500	3547	1	0	20,460	3124	0	1
24,274	3159	1	0	21,419	2752	0	1
27,170	3621	1	0	25,160	3429	0	1
30,168	3782	1	0	22,482	3947	0	0
26,525	4247	1	0	20,969	2509	0	0
27,360	3982	1	0	27,224	5440	0	0
21,690	3568	1	0	25,892	4042	0	0
21,974	3155	1	0	22,644	3402	0	0
20,816	3059	1	0	24,640	2829	0	0
18,095	2967	1	0	22,341	2297	0	0
20,939	3285	1	0	25,610	2932	0	0
22,644	3914	1	0	26,015	3705	0	0
24,624	4517	0	1	25,788	4123	0	0
27,186	4349	0	1	29,132	3608	0	0
33,990	5020	0	1	41,480	8349	0	0
23,382	3594	0	1	25,845	3766	0	0
20,627	2821	0	1				

لاحظ أن : $D_2 = 1$ للولايات في الشمال الشرقي والشمال الأوسط ، 0 بخلاف ذلك .
$D_3 = 1$ للولايات في الجنوب ، 0 بخلاف ذلك .

المصدر : National Educational Association, as reported by Albuquerque Tribune, Nov. 7, 1986.

لاحظ أن (1.2.9) مماثلة لأي نموذج انحدار متعدد سبق وذكرناه، بخلاف أن المتغيرات المنحدر عليها هي متغيرات نوعية وليست كمية ، وبالتالي هي متغيرات وهمية تأخذ القيمة 1 إذا كانت المشاهدة تنتمي إلى فئة ما و 0 إذا كانت لا تنتمي لهذه الفئة . ومن الآن فصاعداً سنرمز للمتغيرات الوهمية بالحرف D. جدول (1.9) يوضح المتغيرات الوهمية التي تم تكوينها .

ماذا يقول لنا النموذج (1.2.9)؟ إذا افترضنا أن حد الخطأ مستوفي فروض الـ OLS التقليدية ، وبإدخال التوقع على طرفي (1.2.9)، نحصل على :

متوسط راتب مدرسي المدارس الحكومية في الشمال الشرقي وشمال الوسط :

$$E(Y_i \mid D_{2i} = 1, D_{3i} = 0) = \beta_1 + \beta_2 \qquad (2.2.9)$$

متوسط راتب مدرسي المدارس الحكومية في الجنوب :

$$E(Y_i \mid D_{2i} = 0 , D_{3i} = 1) = \beta_1 + \beta_3 \qquad (3.2.9)$$

وقد تسأل كيف لنا الآن معرفة متوسط راتب المدرسين في الغرب. إذا ضمنت أنه يساوي B1، ستكون على صواب تماماً.

فمتوسط راتب مدرسي المدارس الحكومية في الغرب :

$$E(Y_i \mid D_{2i} = 0 , D_{3i} = 0) = \beta_1 \qquad (4.2.9)$$

بمعنى آخر، متوسط راتب مدرسي المدارس الحكومية في الغرب يساوي الجزء المقطوع من المحور الصادي β_1 حتى الانحدار المتعدد (1.2.9)، أما معاملات الميل β_2 و β_3 فإنها تعبر عن مدى اختلاف متوسط راتب المدرسين في الشمال الشرقي وشمال الوسط والجنوب عن متوسط راتب المدرسين في الغرب. ولكن كيف يمكن معرفة ماذا كانت هذه الفروق معنوية؟

قبل الإجابة على هذا السؤال، دعنا نستعرض نتائج الانحدار (1.2.9) فباستخدام البيانات المعطاة في الجدول (1.9)، حصلنا على النتائج التالية :

$$\begin{aligned} \hat{Y}_i &= 26{,}158.62 \quad - \ 1734.473 D_{2i} - \ 3264.615 D_{3i} \\ se &= (1128.523) \quad (1435.953) \quad (1499.615) \\ t &= (23.1759) \quad (-1.2078) \quad (-2.1776) \\ & \quad (0.0000)^* \quad (0.2330)^* \quad (0.0349)^* \quad R^2 = 0.0901 \end{aligned} \qquad (5.2.9)$$

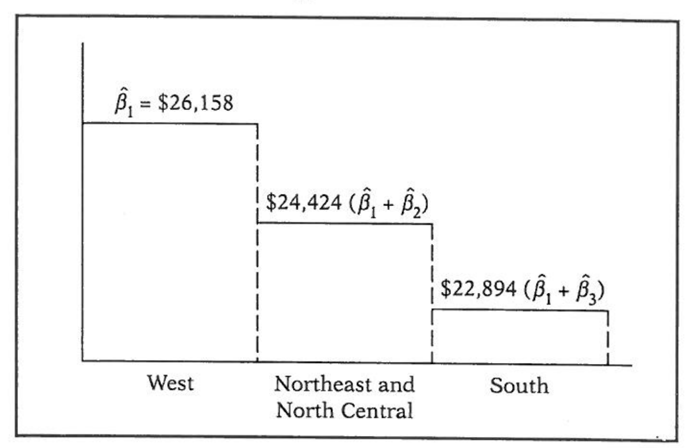
حيث : * تعني قيم الـ p-value.

كما توضح نتائج الانحدار، متوسط رواتب المدرسين في الغرب حوالي 26,158$ أما المدرسون في الشمال الشرقي وشمال الوسط فأقل بحوالي 1734$، والمدرسون في الجنوب أقل بحوالي 3256$. متوسط الرواتب الحقيقية في المنطقتين الأخيرتين يمكن الحصول عليهما بسهولة بإضافة هذه الفروق في الرواتب إلى متوسط راتب المدرسين في الغرب، كما هو موضح في المعادلتين (3.2.9) و(4.2.9). وإذا قمنا بذلك، سنجد أن متوسط الرواتب في المنطقتين الأخيرتين هو تقريباً 24,242$ و 22,894$.

كيف يمكنك معرفة أن هذه المتوسطات في الرواتب مختلفة إحصائياً عن متوسط الرواتب في الغرب، فئة المقارنة؟ هذه السؤال سهل نسبياً. فكل ما نحتاج فعله هو معرفة ما إذا كانت معاملات الميل في (5.2.9) معنوية إحصائياً. كما هو واضح من الانحدار، معامل الميل المقدر للمثال الشرقي وشمال الوسط غير معنوي إحصائياً. حيث إن قيمة p-value هي 3,5% فقط. وبالتالي الاستنتاج العام أنه إحصائياً متوسط الراتب للمدرسين بالمدارس الحكومية في الغرب والشمال الشرقي وشمال الوسط

تقريباً متساويين ، ولكن متوسط الراتب للمدرسين في الجنوب أقل معنوياً بحوالي 3265$. بالرسوم، هذا الوضع موضح في الشكل (1.9).

المهم الآن هو مراعاة الترتيب في تفسير هذه الفروق. المتغيرات الوهمية ستوضح هذه الفروق، إذا كانت موجودة، ولكن لا تقترح أسبابًا لهذه الفروق.



الشكل (1.9) متوسط الرواتب (بالدولار) للمدرسين العاملين في المدارس الحكومية في ثلاث مناطق مختلفة

الفروق في مستويات التعليم، مؤشرات تكلفة المعيشة والنوع والعرق ممكن يكون لها جميعاً بعض التأثير على الفروق المشاهدة. وبالتالي إذا لم نضع في الاعتبار كل المتغيرات التي يمكن أن تؤثر على راتب المدرس، لن تستطيع تحديد مسببات هذه الفروق.

من المناقشة السابقة، واضح أنه يجب على الباحث أن يرى ما إذا كانت المعاملات المرتبطة بالمتغيرات الوهمية المختلفة لها معنوية إحصائية كل على حدة أم لا. وهذا المثال يوضح أيضاً كم هو سهل التعامل مع المتغيرات المنحدرة عليها عندما تكون وهمية أو نوعية في نماذج الانحدار.

محاذير استخدام المتغيرات الوهمية: Caution in the Use of Dummy Variables

على الرغم من سهولة إدخال المتغيرات الوهمية في نماذج الانحدار ، لابد أن يستخدمها الفرد بحذر شديد. **وبالتحديد دعنا نستعرض النقاط التالية:**

1 – في مثال 1.9 ، للتمييز بين المناطق الثلاث ، استخدمنا متغيرين وهميين فقط وهما D_2 و D_3. لماذا لم نستخدم ثلاثة متغيرات وهمية للتعبير عن الثلاث مناطق؟ دعنا نفترض أننا قمنا بذلك ، وأصبح لدينا النموذج (1.2.9) كالتالي:

$$Y_i = \alpha + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \qquad (6.2.9)$$

حيث : D_{1i} تأخذ القيمة 1 للولايات الموجودة في الغرب، و 0 بخلاف ذلك.

الآن لدينا متغير وهمي لكل منطقة من المناطق الجغرافية الثلاث.

باستخدام البيانات الموجودة في جدول (1.9)، إذا قمت بعمل الانحدار (6.2.9) سيرفض جهاز الحاسب نموذج الانحدار (جرب ذلك)(6) لماذا؟. السبب هو أن تكوين (6.2.9) عندما يكون لدينا متغير وهمي لكل مجموعة أو فئة، بالإضافة إلى الثابت الممثل للجزء المقطوع من المحور الصادي، يكون لدينا حالة من الارتباط التام، أي وجود علاقة خطية تامة بين المتغيرات لماذا؟ بالرجوع إلى الجدول (1.9) تخيل أننا أضفنا عمود D_1 والذي يأخذ القيمة 1 عندما تكون الولاية في الغرب، و 0 بخلاف ذلك. إذا أضفنا الآن ثلاثة أعمدة أخرى بشكل أفقي، سنحصل على عمود يحتوي على 51 قيمة تساوي كل منها الواحد الصحيح، ولكن بما أن قيمة الجزء المقطوع من المحور الصادي α تساوي الواحد لكل مفردة، سيكون لديك عمود يحتوي أيضاً على 51 قيمة تساوي كل منها الواحد الصحيح، بمعنى آخر، مجموعة الثلاثة أعمدة D سينتج عنها في النهاية عمود الجزء المقطوع من المحور الصادي، وبالتالي نصل إلى الارتباط التام. في هذه الحالة، تقدير النموذج (6.2.9) يصبح مستحيلاً.

ما نستنتج من ذلك هو: إذا كان المتغير النوعي له m طبقة تستخدم فقط $(m-1)$ من المتغيرات الوهمية. في مثالنا الحالي، بما أن المتغير النوعي «المنطقة الجغرافية» له ثلاث طبقات تستخدم اثنين من المتغيرات الوهمية. إذا لم تتبع هذه القاعدة، ستنتج فيما يسمى مصيدة المتغير الوهمي، أي وجود ارتباط تام أو ارتباط تام متعدد، إذا كان لدينا أكثر من علاقة تام بين المتغيرات.

هذه القاعدة تطبق أيضاً إذا كان لدينا أكثر من متغير نوعي واحد في النموذج، وسنقدم مثالاً على ذلك لاحقًا. ولذلك يجب أن نحدد القاعدة السابقة كالتالي: لكل متغير منحدر عليه نوعي، عدد المتغيرات الوهمية التي يمكننا استخدامها يجب أن تكون أقل بواحد من عدد طبقات هذا المتغير. إذا

(6) ستحصل على رسالة مكتوب فيها أن مصفوفة البيانات منفردة.

كان لدينا مثلاً في المثال 1.9 معلومات عن نوع المدرس، لابد من استخدام متغير وهمي إضافي (متغير واحد وليس اثنين) يأخذ القيمة 1 للإناث و 0 للذكور أو العكس.

2 - الطبقة التي لا يوجد متغير وهمي يمثلها تعرف باسم طبقة الأساس، طبقة التحكم أو المقارنة أو الطبقة المرجعية أو المحذوفة. وكل المقارنات تتم من خلال العلاقة مع هذه الطبقة.

3 - قيمة الجزء المقطوع من المحور الصادي (β_1) يمثل قيمة متوسط الطبقة المرجعية. في مثال 1.9، الطبقة المرجعية هي ا لمنطقة الغربية. وبالتالي، في الانحدار (5.2.9) قيمة الجزء المقطوع من المحور الصادي تساوي تقريباً 26,159 تمثل متوسط الراتب للمدرسين في الولايات الغربية.

4 - المعاملات المرتبطة بالمتغيرات الوهمية في (1.2.9) معروفة باسم معاملات الجزء المقطوع من المحور الصادي التفاضلية، حيث إنها تعبر عن التغير في الجزء المقطوع من المحور الصادي عند تغيير معامل الجزء المقطوع من المحور الصادي للطبقة المرجعية بواحد صحيح.

فعلى سبيل المثال، في (5.2.9) القيمة 1734- تعني أن متوسط رواتب المدرسين في الشمال الشرقي وشمال الوسط أقل بحوالي 1734$ من متوسط رواتب نظيرهم المساوي لـ 26,159$ في الطبقة المرجعية، وهي الغرب.

5 - إذا كان المتغير النوعي له أكثر من طبقة واحدة، كما في مثالنا التوضيحي، اختبار الطبقة المرجعية يرجع تماماً للباحث، أحياناً اختبار الطبقة المرجعية يكون راجعاً لمشكلة ما محددة.

في مثالنا التوضيحي، ممكن أن نختار الجنوب كطبقة مرجعية، في هذه الحالة، نتائج الانحدار المعطاة في (5.2.9) ستتغير، حيث إن كل المقارنات ستتم من خلال العلاقة مع الجنوب. بالطبع ذلك لن يغير الاستنتاجات الكلية (لماذا؟). في هذه الحالة، قيمة الجزء المقطوع من المحور الصادي ستكون تقريباً 22,894$ والتي تمثل متوسط رواتب المدرسين في الجنوب.

6 - سبق وحذرنا من قبل من مصيدة المتغيرات الوهمية. هناك طريقة للهروب من هذه المصيدة، بتقديم عدد من المتغيرات الوهمية مساو لعدد طبقات المتغير

النوعي مع عدم إدخال جزء مقطوع من المحور الصادي في هذا النموذج . وبالتالي إذا حذفنا الجزء المقطوع من المحور الصادي من (6.2.9) سيكون لدينا النموذج التالي :

$$Y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + u_i \qquad (7.2.9)$$

في هذه الحالة ، نكون نجونا من مصيدة التغير الوهمي ، حيث إنه لم يعد هناك ارتباط تام . ولكن تأكد عند إجرائك لهذا الانحدار على الحاسب الآلي أنك تختار حالة الانحدار بدون جزء ثابت مقطوع من المحور الصادي .

كيف يمكننا تفسير الانحدار (7.2.9)؟ إذا أدخلنا التوقع على (7.2.9) **سنحصل على :**

β_1 = متوسط رواتب المدرسين في الغرب .

β_2 = متوسط رواتب المدرسين في الشمال الشرقي وشمال الوسط .

β_3 = متوسط رواتب المدرسين في الجنوب .

بمعنى آخر ، في حالة عدم وجود جزء مقطوع من المحور الصادي ، والسماح للمتغير الوهمي بأن يعبر عن كل طبقة ، نحصل مباشرة على القيم المتوسطة للطبقات المختلفة . **نتائج (7.2.9) والخاصة بمثالنا التوضيحي معطاة كالتالي :**

$$\begin{aligned} \hat{Y}_i &= 26{,}158.62 D_{1i} + 24{,}424.14 D_{2i} + 22{,}894 D_{3i} \\ \text{se} &= (1128.523) \quad (887.9170) \quad (986.8645) \\ t &= (23.1795)^{*} \quad (27.5072)^{*} \quad (23.1987)^{*} \\ & R^2 = 0.0901 \end{aligned} \qquad (8.2.9)$$

حيث : * تعبر عن الـ p-valus والخاصة بنسب الـ t وهي صغيرة جداً .

كما ترى ، المعاملات الوهمية تعطي مباشرة متوسط (الراتب) في المناطق الثلاث ، الغرب ، الشمال الشرقي وشمال الوسط والجنوب .

7- أي طريقة تعتبر الأفضل عند استخدام المتغيرات الوهمية : (1) التعامل مع متغير وهمي لكل طبقة وحذف الجزء الثابت المقطوع من المحور الصادي أو (2) وجود جزء ثابت في النموذج و استخدام $(m-1)$ فقط من المتغيرات الوهمية ، حيث m تمثل عدد طبقات المتغير النوعي؟ حاول Kennedy الإجابة عند ذلك كالتالي :

العديد من الباحثين وجدوا أن المعادلة مع جزء ثابت من المحور الصادي أفضل ، حيث تسمح له بالإجابة عن السؤال الأكثر أهمية وهو بالتحديد ما إذا كان هناك فروق بين الطبقات أم لا ؟ وإذا كان هناك فرق فما هي قيمة هذا الفرق . إذا كانت

هناك فروق بين الطبقات، يتم قياسها مباشرة من خلال تقدير معامل المتغير الوهمي. اختيار ما إذا كان تقسيم ذلك المتغير النوعي إلى طبقات شيء مهم أو لا، يمكن إجراؤه من خلال اختيار t لمعامل المتغير الوهمي، ويكون الاختيار حول الصفر (أو بشكل أكثر عمومية، اختيار F لمجموعة من تقديرات معاملات المتغيرات الوهمية)(7).

3.9 نماذج ANOVA لاثنين من المتغيرات النوعية: ANOVA MODELS WITH TWO QUALITATIVE VARIABLES

في الفقرة السابقة، استعرضنا نموذج ANOVA بمتغير نوعي واحد فقط له ثلاث طبقات. في هذه الفقرة سنستعرض نموذج ANOVA آخر لديه اثنان من المتغيرات النوعية، مما يجعلنا نناقش بعض النقاط الإضافية والخاصة بالمتغيرات الوهمية.

مثال 2.9

الأجر بالساعة وعلاقته بالحالة الاجتماعية ومحل الإقامة :

HOURLY WAGES IN RELATION TO MARITAL STATUS AND REGION OF RESIDENCE

من خلال بيانات عينة من 528 شخصًا في مايو 1985. تم الحصول على نتائج الانحدار التالية :(8)

$$\begin{aligned} \hat{Y}_i &= 8.8148 + 1.0997D_{2i} - 1.6729D_{3i} \\ se &= (0.4015) \quad (0.4642) \quad (0.4854) \\ t &= (21.9528) \quad (2.3688) \quad (-3.4462) \\ & \quad (0.0000)^* \quad (0.0182)^* \quad (0.0006)^* \\ & R^2 = 0.0322 \end{aligned} \tag{1.3.9}$$

حيث : Y = الأجر بالساعة ($)

D_2 = الحالة الاجتماعية، 1 = متزوج، 0 = بخلاف ذلك.

D_3 = محل الإقامة، 1 = الجنوب، 0 = بخلاف ذلك.

و * ترمز إلى الـ p-value.

في هذا المثال، لدينا متغيران نوعيان كل منهما له طبقتان. وبالتالي فهناك متغير وهمي لكل طبقة.

---

(7) Peter Kennedy, A Guide to Econometrics, 4th ed., MIT Press, Cambridge, Mass., 1998, p. 223.

(8) البيانات تم الحصول عليها من بنك بيانات موجود في Arthur S. Goldberger, Introductory Econometrics, Harvard University Press, Cambridge, Mass., 1998. قمنا باستخدام هذه البيانات من قبل في الفصل (2).

أي طبقة تعتبر المرجعية هنا؟ واضح أنها الشخص غير المتزوج ومحل الإقامة في أي مكان غير الجنوب. بمعنى آخر، الأشخاص غير المتزوجين والذين لا يعيشون في الجنوب هم الفئة المحذوفة. وبالتالي كل المقارنات ستتم بالمقارنة مع هذه الطبقة. متوسط الأجر بالساعة في هذه الطبقة هو 8,81$، بالمقارنة مع ذلك، متوسط الأجر بالساعة للمتزوجين أعلى بحوالي 1,1$. لحساب متوسط الأجر الحقيقي هو 9,91$ (= 8,81 + 1,10).

بالمقارنة مع المقيمين في الجنوب، متوسط الأجر بالساعة أقل بحوالي 1,67$ وبالتالي متوسط الأجر الحقيقي هو 7,14$.

هل متوسطات الأجور السابقة مختلفة إحصائياً عن طبقة الأساس؟ نعم مختلفون حيث إن قيمها الخاصة بالـ p-value منخفضة بشكل كاف.

الملاحظة المهمة هنا هي: طالما لديك أكثر من متغير نوعي واحد لابد من الاهتمام بالطبقة التي يتم استخدامها كطبقة أساساً. حيث إن كل المقارنات ستتم من خلال هذه الطبقة. هذه النقطة مهمة للغاية عندما يكون لديك متغيرات نوعية عديدة، كل منها له عدة طبقات. ولكن الأسلوب الذي يتم به إدخال العديد من المتغيرات النوعية يجب أن يكون قد اتضح الآن بعد هذه المناقشة.

4.9 الانحدار بمزيج من المتغيرات المنحدر عليه النوعية والكمية: نماذج ANCOVA

REGRESSION WITH A MIXTURE OF QUANTITATIVE AND QUALITATIVE REGRESSORS: THE ANCOVA MODELS

نماذج ANOVA التي تمت مناقشتها في الفقرتين السابقتين، على الرغم من كثرة استخدامها في العلوم الاجتماعية و النفسية والتعليمية وأبحاث السوق، إلا أنها غير متعارف عليها في الاقتصاد. بالأخص في العديد من الأبحاث الاقتصادية، نموذج الانحدار يحتوي على بعض المتغيرات المفسرة النوعية، وبعضها الآخر متغيرات مفسرة كمية نماذج الانحدار التي تحتوي على مزيج من المتغيرات النوعية و الكمية تسمى نماذج تحليل التغاير (Analysis of covariance Models (ANCOVA .

نماذج ANCOVA هي امتداد لنماذج الـ ANOVA، حيث إنها تعتبر طريقة للتحكم الإحصائي في أثر المتغيرات المنحدرة عليها النوعية، وتسمى متغيرات التحكم (أو المتغيرات المعاونة) في نموذج يحتوي على كل من المتغيرات المفسرة النوعية والكمية.

دعنا الآن نشرح نماذج الـ ANCOVA.

لتسهيل عملية التحليل، دعنا نعتبر مرة أخرى مثال 1.9 مع اعتبار أن متوسط الرواتب للمدرسين بالمدارس الحكومية قد لا يختلف في الثلاث مناطق الجغرافية المذكورة سابقاً، وذلك إذا أخذنا في الاعتبار أي متغيرات أخرى غير موحدة على كل هذه المناطق.

دعنا نعتبر على سبيل المثال متغير النفقات الخاصة بالمدارس الحكومية من السلطات المحلية، حيث إن التعليم الحكومي مسئولية السلطة المحلية أو سلطة الولاية. لندرس تلك الحالة دعنا نستعرض النموذج التالي:

$$Y_i = \beta_1 + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 X_i + u_i \qquad (1.4.9)$$

حيث: Y_i = متوسط الراتب السنوي للمدرسين في المدارس الحكومية في الولاية ($).

X_i = الإنفاق على المدارس الحكومية ($) بالنسبة للفرد.

D_{2i} = 1، إذا كانت الولاية في الشمال الشرقي أو شمال الوسط

= 0، بخلاف ذلك.

D_{3i} = 1، إذا كانت الولاية في الجنوب.

= 0، بخلاف ذلك.

بيانات المتغير X معطاة في الجدول (1.9). ضع في الاعتبار أننا نعامل الغرب على أنه الطبقة المرجعية. وأيضاً لاحظ أنه بالإضافة إلى المتغيرين النوعيين اللذين يمثلان متغيرين نوعيين لدينا متغير آخر كمي وهو X والذي يسمى في إطار الـ ANCOVA على أنه متغير معاون COVARIATE، كما سبق وذكرنا من قبل.

مثال 3.9

رواتب المدرسين بالنسبة للفرد وعلاقتها بالمنطقة والإنفاق على المدارس الحكومية:

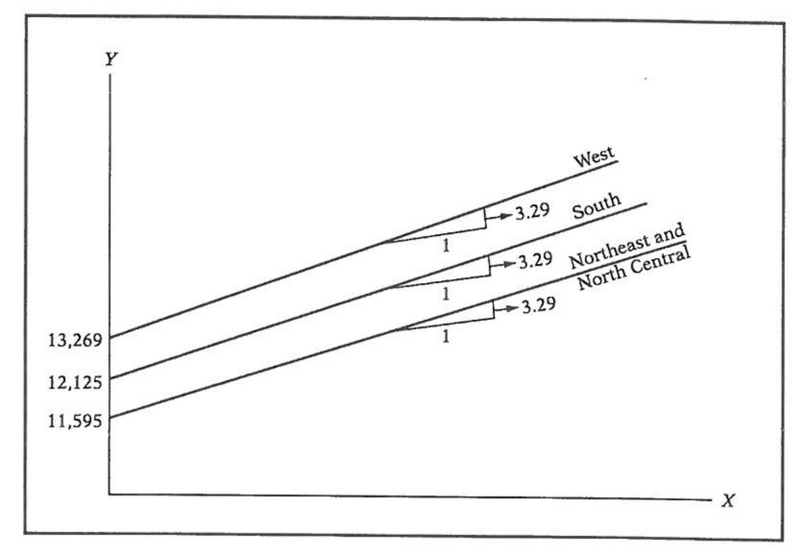
TEACHER'S SALARY IN RELATION TO REGION AND SPENDING ON PUBLIC SCHOOL PER PUPIL

من البيانات المعطاة في جدول (1.9)، نتائج نموذج (1.4.9) معطاة كالتالي:

$$\begin{aligned} \hat{Y}_i &= 13{,}269.11 - 1673.514 D_{2i} - 1144.157 D_{3i} + 3.2889 X_i \\ se &= (1395.056) \quad (801.1703) \quad (861.1182) \quad (0.3176) \qquad (2.4.9) \\ t &= (9.5115)^{*} \quad (-2.0889)^{*} \quad (-1.3286)^{**} \quad (10.3539)^{*} \\ & \qquad R^2 = 0.7266 \end{aligned}$$

حيث : * تعني قيم الـ p-value أقل من 5%، و** تعني أن قيم الـ p-values أكبر من 5% وكما نرى من النتائج، كلما زاد متوسط الإنفاق الحكومي بدولار واحد، فإن راتب المدرس العامل بالمدرسة الحكومية يزداد بحوالي 3,29$. الآن إذا افترضنا ثبات الإنفاق على التعليم نرى أن معامل الجزء الثابت من المحور الصادي التفاضلي معنوي عند منطقة الشمال الشرقي وشمال الوسط وغير معنوي عند الجنوب. هذه النتائج مختلفة عن التي حصلنا عليها من قبل في (5.2.9). ولكن هذا يجب ألا يكون مفاجأة، حيث إننا في (5.2.9) لم نضع في الاعتبار المتغير المساعد، وهي الفروق في الإنفاق على التعليم بالنسبة للفرد. ولتوضيح ذلك بيانياً لدينا الشكل (2.9).

لاحظ أنه على الرغم من أننا وضحنا ثلاث خطوات انحدار للثلاث مناطق، فإنه إحصائياً تكون خطوط الانحدار متماثلة في الغرب والجنوب. ولاحظ أيضاً أن خطوط الانحدار الثلاثة رسمت متوازية (لماذا؟).



الشكل (2.9)- راتب مدرس المدرسة الحكومية (Y) وعلاقته بالإنفاق على التعليم بالنسبة للفرد (X)

Puldic school teacher's salary (Y) in relation to per pupil Expenotitine on education (X)

5.9 المتغير الوهمي كبديل لاختبار CHOW :
THE DUMMY VARIABLE ALTERNATIVE TO THE CHOW TEST

في الفقرة 8.8 ناقشنا اختبار Chow، والذي يستخدم لاختبار الاستقرار الهيكلي لنموذج الانحدار. المثال الذي ناقشناه كان مرتبطًا بالعلاقة بين الادخار والدخل في الولايات المتحدة خلال الفترة 1970– 1995. قمنا بتقسيم فترة العينة إلى جزءين :

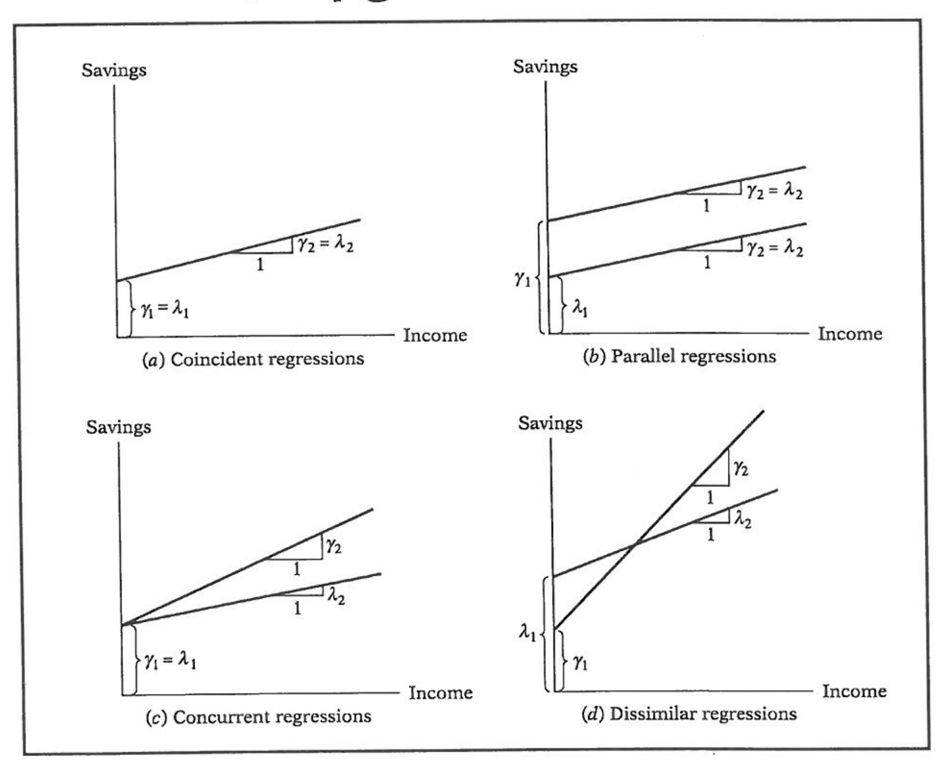
1965 – 1981، 1982 – 1995، ووضحنا أنه باستخدام اختبار Chow هناك فرق بين الانحدار الخاص بالادخار على الدخل في الفترتين السابقتين.

عموماً لا نستطيع ما إذا كان الفرق بين الانحدارين يرجع إلى الفرق بين حدود الأجزاء الثابتة المقطوعة من المحور الصادي أم معاملات الميل أم الاثنين معاً. هذه المعلومة في حد ذاتها تعتبر معلومة في غاية الأهمية.

بالعودة إلى المعادلتين (1.8.8) و(2.8.8)، نرى أن هناك أربعة احتمالات، التي نستعرضها أيضاً في الشكل (3.9)، **هذه الاحتمالات هي**:

1 – كل من معاملات الجزء الثابت والميل متساوية في الانحدارين. وهذا يسمى الانحدارات المتطابقة، كما هو موضح في الشكل (3.9).

2 – الجزء المقطوع من المحور الصادي مختلف في الانحدارين، ولكن الميل متساو، وهذا يسمى الانحدارات المتوازية، وهو موضح في الشكل (3.9).



شكل (3.9) الانحدارات الخاصة بالادخار – الدخل موضح بالرسم

Plausible savings - income regressions

3 - الجزء المقطوع من المحور الصادي متساو في الانحدارين ، ولكن الميل مختلف ، وهذا يسمى الانحدارات المتزامنة (شكل 3c.9) .

4 - كل من الأجزاء المقطوعة، من المحور الصادي والميول غير متساوية . ويسمى هذا الانحدارات غير المتماثلة، وهو موضح في الشكل (3.9) .

اختبار Chow متعدد الخطوات الذي ذكرناه من قبل في الفقرة 8.8، وكما سبق وأوضحنا، يخبرنا فقط ما إذا كان الانحداران (أو أكثر) مختلفين أم لا، ولكنه لا يتطرق إلى مصدر هذا الاختلاف. مصدر الاختلاف، إذا كان هناك اختلاف، يمكن تحديده عن طريق تجميع كل المشاهدات (كل الـ 26 مفردة) وإجراء انحدار متعدد كالموضح أسفل :[10]

$$Y_t = \alpha_1 + \alpha_2 D_t + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \qquad (1.5.9)$$

حيث : Y = الادخار

X = الدخل

t = الزمن

D = 1 إذا كانت المفردة واقعة في الفترة 1982 – 1995 .

= 0 بخلاف ذلك (أي أن المفردة واقعة في الفترة 1970 – 1981) .

جدول (2.9) بيانات الادخار والدخل في الولايات المتحدة ، **1970–1995**

Saering and Iacove Data, Crnted states, 1970–1995

Observation	Savings	Income	Dum
1970	61	727.1	0
1971	68.6	790.2	0
1972	63.6	855.3	0
1973	89.6	965	0
1974	97.6	1054.2	0
1975	104.4	1159.2	0
1976	96.4	1273	0
1977	92.5	1401.4	0
1978	112.6	1580.1	0
1979	130.1	1769.5	0
1980	161.8	1973.3	0
1981	199.1	2200.2	0
1982	205.5	2347.3	1
1983	167	2522.4	1
1984	235.7	2810	1

(10) كما في اختبار Chow ،الأسلوب التجميعي يفترض تساوي التباين، أي أن $\sigma^2 = \sigma_2^2 = \sigma_1^2$.

1985	206.2	3002	1
1986	196.5	3187.6	1
1987	168.4	3363.1	1
1988	189.1	3640.8	1
1989	187.8	3894.5	1
1990	208.7	4166.8	1
1991	246.4	4343.7	1
1992	272.6	4613.7	1
1993	214.4	4790.2	1
1994	189.4	5021.7	1
1995	249.3	5320.8	1

لاحظ أن: Dum = 1 لكل المشاهدات الو اقعة في 1982، 0 بخلاف ذلك، الادخار والدخل مقاسان بالبليون دولار

المصدر: Elanonic Report of the presidert, 1997, table B-28, p.332.

جدول (2.9) يوضح الهيكل الخاص بمصفوفة البيانات .

لتوضيح توابع (1.5.9) مع الافتراض التقليدي بأن $E(u_i) = 0$ نحصل على التالي :

دالة الادخار المتوسطة للفترة 1970 – 1981

$$E(Y_t \mid D_t = 0, X_t) = \alpha_1 + \beta_1 X_t \qquad (2.5.9)$$

دالة الادخار المتوسطة للفترة 1982 – 1995

$$E(Y_t \mid D_t = 1, X_t) = (\alpha_1 + \alpha_2) + (\beta_1 + \beta_2) X_t \qquad (3.5.9)$$

يمكن للقارئ ملاحظة أن هذه الدوال مماثلة للدوال المذكورة في (1.8.8) و(2.8.8) مع استخدام $\alpha_1 = \lambda_1$، $\beta_2 = \lambda_2$، $(\alpha_1 + \alpha_2) = \gamma_1$ و $(\beta_1 + \beta_2) = \gamma_2$ وبالتالي تقرير (1.5.9) مساو لتقدير دالتي الادخار المنفصلتين (1.8.8) و(2.8.8) .

في (1.5.9)، α_2 تسمى الجزء الثابت والمقطوع من المحور الصادي التفاضلي كما كان في السابق، وβ_2 يسمى معامل الميل التفاضلي (ويسمى أيضاً الميل الثابت) وهو يوضح كم التغيير في معامل الميل الخاص بدالة الادخار الثانية عن نظيره والخاص بالدالة الأولى (المرحلة الثانية هي الطبقة التي يأخذ فيها المتغير الوهمي القيمة 1) . لاحظ كيف أن إدخال المتغير الوهمي D بصورة تفاعلية أو ضربية (D مضروبة في X) ساعدنا على التفريق بين معاملات الميل للفترتين المختلفتين، وذلك مثلما كان إدخال المتغير الوهمي بشكل تجميعي ساعدنا على التفرقة بين الجزء المقطوع من المحور الصادي في الفترتين المختلفتين .

مثال 4.9

الفروق الهيكلية في انحدار الادخار - الدخل في الولايات المتحدة باستخدام أسلوب المتغير الوهمي :

STRUCTURAL DIFFERENCES IN THE U.S. SAVINGS - INCOME REGRESSION, THE DUMMY VARIABLE APPROACH

قبل أن نستعرض مثالنا الحالي، دعنا أولاً نقدم نتائج نموذج (1.5.9) بعد تطبيقه على بيانات الادخار - الدخل الخاص بالولايات المتحدة.

$$\begin{aligned}\hat{Y}_t &= 1.0161 + 152.4786D_t + 0.0803X_t - 0.0655(D_tX_t)\\ \text{se} &= (20.1648)\quad (33.0824)\quad (0.0144)\quad (0.0159)\\ t &= (0.0504)^{**}\quad (4.6090)^{*}\quad (5.5413)^{*}\quad (-4.0963)^{*}\end{aligned} \tag{4.5.9}$$

$$R^2 = 0.8819$$

حيث : * تعني أن الـ p-value أقل من 5% و** تعني أن p-value أكثر من 5% كما يتضح من نتائج الانحدار، كل من الجزء الثابت التكافلي ومعاملات الميل لها معنوية إحصائية، مما يعني أن هناك اختلافًا في انحدار الادخار - الدخل في الفترتين السابقتين وهذا موضح في الشكل (3.9).

من (4.5.9)، يمكننا اشتقاق المعادلتين (2.5.9) و(3.5.9) وهما كالتالي :

انحدار الادخار - الدخل، 1970 - 1981

$$\hat{Y}_t = 1.0161 + 0.0803X_t \tag{5.5.9}$$

انحدار الادخار - الدخل، 1982 - 1995

$$\begin{aligned}\hat{Y}_t &= (1.0161 + 152.4786) + (0.0803 - 0.0655)X_t\\ &= 153.4747 + 0.0148X_t\end{aligned} \tag{6.5.9}$$

وهذه النتائج متطابقة تماماً مع النتائج التي سبق وحصلنا عليها في (1a.8.8) و(2a.8.8) وقد كان ذلك أمراً متوقعًا. هذه الانحدارات سبق وتم توضيحها في الشكل (3.8).

فائدة استخدام أسلوب المتغير الوهمي [أي تقدير (1.5.9)] مقارنةً باختبار Chow [أي تقدير الانحدارات الثلاثة (1.8.8)، (2.8.8) و(3.8.8)] يمكن تحديدها كالتالي :

1 - نحن نحتاج لإجراء انحدار واحد فقط، حيث إن الانحدارات الجزئية يمكن استنتاجها بسهولة كما فعلنا في المعادلتين (2.5.9) و(3.5.9).

2 - الانحدار الوحيد (1.5.9) يمكن استخدامه لإجراء العديد من اختبارات الفروض. فمثلاً إذا كان معامل الجزء الثابت التفاضلي α_2 ليس له معنوية إحصائية يمكن أن

تقبل الفرض القائل بأن الانحدارين الآتيين لهما نفس الجزء الثابت المقطوع من المحور الصادي، أي أن الانحدارين متزامنين (انظر الشكل 3.9). بالمثل، إذا كان معامل الميل التفاضلي β_2 ليس له معنوية إحصائية ولكن 2a معنوي، قد لا نرض الفرض القائل بأن الانحدارين لهما نفس الميل، أي أن خطي الانحدارين متوازيان (انظر الشكل 3b.9). إختبار استقرار الانحدار ككل (أي أن $\alpha_2 = \beta_2 = 0$ آنياً) يمكن إجراؤه عن طريق اختبار F التقليدي (تذكر اختبار F للمربعات الصغرى المقيدة) إذا لم نرفض الفرض ستكون خطوط الانحدار متطابقة كما هو موضح في الشكل (3.9).

3 - اختبار Chow لا يوضح لنا بشكل صريح أي معامل مختلف سواء كان الجزء الثابت أو الميل أو كما في (مثالنا الحالي) كل منهما مختلف في الفترتين. أي أن الباحث يمكن أن يحصل على اختبار Chow معنوي، لأن الميل مختلف أو الجزء الثابت يكون هو المختلف أو كلاهما معاً. بمعنى آخر، لا نستطيع تحديد، باستخدام اختبار Chow، أي من الاحتمالات الأربعة المذكورة في الشكل (2.9) هي المتحققة. ولهذا السبب فإن أسلوب المتغير الوهمي له ميزة خاصة وفريدة فهو لا يحدد لنا فقط ما إذا كان هناك اختلاف، ولكن يحدد أيضاً مصدر هذا الاختلاف - تستطيع معرفة ما إذا كان ذلك راجعًا للميل أو راجعًا إلى الجزء الثابت المقطوع من المحور الصادي أو إلى كليهما معاً.

عند التطبيق العملي تكون معرفة مصدر الاختلاف بين الانحدارين لها نفس الأهمية أو قد تزيد من الأهمية عن المعرفة الخاصة فقط بوجود اختلاف من عدمه بين الانحدارين.

4 - في النهاية، بما أن التجميع (أي استخدام كل المفردات في انحدار واحد) يزيد من درجات الحرية، فقد يزيد من الدقة النسبية لتقدير المعاملات. بالطبع يجب أن نضع في الاعتبار أن كل إضافة لمتغير وهمي ستستهلك درجة من درجات الحرية.

6.9 الأثر التفاعلي باستخدام المتغيرات الوهمية:
INTERACTION EFFECTS USING DUMMY VARIABLES

المتغيرات الوهمية هي أداة مرنة لدراسة العديد من المشكلات. لنرى ذلك دعنا نعتبر النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta X_i + u_i \qquad (1.6.9)$$

حيث: Y = الأجر بالساعة بالدولار

X = التعليم (عدد سنوات الالتحاق بالمدرسة)

$D_2 = 1$ للأنثى، 0 بخلاف ذلك

$D_3 = 1$ للأشخاص ذوى البشرة غير البيضاء وغير المنتمين لأمريكا واللاتينيين، 0 بخلاف ذلك

في هذا النموذج، النوع والعرق متغيران منحدر عليهما نوعية، والتعليم متغير منحدر عليه كمي[11]. وهذا النموذج يفترض فيه صراحة أن التأثير التفاضلي للمتغير الوهمي الخاص بالنوع D_2 ثابت على مستوى طبيعتي متغير العرق، والتأثير التفاضلي للمتغير الوهمي للعرق D_3 ثابت أيضاً على مستوى طبعتي النوع. أي أنه بغض النظر عن العرق (أى سواء كان الأشخاص ينتمون لبشرة غير بيضاء وغير منتمين لأمريكا اللاتينية) يكون متوسط الراتب أعلى للذكور غير الإناث. وبالمثل إذا قلنا إن الأشخاص ذوى البشرة غير البيضاء وغير منتمين لأمريكا اللاتينية لهم أجر أقل من المتوسط يكون ذلك صحيحاً بغض النظر عن النوع، أي بغض النظر عن كونهم إناثًا أو ذكوراً.

في العديد من التطبيقات يكون هذا الفرض غير واقعي. فالمرأة ذات البشرة الداكنة قد يكون أجرها أقل من الرجل الذي له نفس لون البشرة. بعبارة أخرى، هناك تفاعل بين المتغيرين النوعيين D_2 و D_3، وبالتالي تأثيرهما على المتوسط Y ممكن أن يكون تجميعيًا كما في (1.6.9)، أو تأثيراً مضروبًا كما في النموذج التالي:

$$\hat{Y}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 (D_{2i} D_{3i}) + \beta X_i + u_i \qquad (2.6.9)$$

حيث إن المتغيرات معرفة كما في النموذج (1.6.9).

من (2.6.9)، نحصل على:

$$E(Y_i \mid D_{2i} = 1, D_{3i} = 1, X_i) = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4) + \beta X_i \qquad (3.6.9)$$

والمقدار السابق يمثل دالة متوسط الأجر بالساعة للإناث ذوات البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية. لاحظ التالي:

α_2 = الأثر التفاضلي للفرد أنثى.

α_3 = الأثر التفاضلي للفرد ذي البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية.

(11) إذا عرفنا متغير التعليم على أنه: أقل من التعليم الثانوي، تعليم ثانوي، أعلى من التعليم الثانوي، نستخدم في هذه الحالة متغيرين وهميين للتعبير عن هذه الطبقات الثلاث.

α_4 = الأثر التفاضلي للأنثى ذات البشرة غير البيضاء وغير المنتمية لأمريكا اللاتينية.

وهذا يوضح أن متوسط الأجر بالساعة للأنثى ذات البشرة غير البيضاء وغير المنتمية لأمريكا اللاتينية مختلف (بمقدار α_4) عن متوسط الأجر بالساعة عن الإناث بوجه عام، أو عن ذوي البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية. فعلى سبيل المثال، إذا افترضنا أن كل ا لمعاملات الوهمية التفاضلية الثلاثة سالبة، فإن ذلك يعني أن الإناث ذوات البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية لهن متوسط أجر بالساعة، أقل من الإناث أو الأشخاص ذوي البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية، وذلك مقارنة بطبقة الأساس، والتي تتمثل في مثالنا الحالي في الذكور ذوي البشرة البيضاء أو المنتمين لأمريكا اللاتينية.

والآن يمكن للقارئ معرفة أثر المتغير الوهمي التفاعلي (والذي يتمثل في حاصل ضرب اثنين من المتغيرات الوهمية أو النوعية) في توضيح أثر التفاعل بين المتغيرات الوهمية التي تم من قبل دراستها منفردة. (أو بشكل تجميعي).

مثال 5.9

متوسط الأجر المكتسب بالساعة وعلاقته بالتعليم ، النوع والعرق :

AVERAGE HOURLY EARNINGS IN RELATION TO EDUCATION, GENDER AND RACE

دعنا في البداية نستعرض نتائج الانحدار الخاص بالنموذج (1.6.9). فباستخدام البيانات التي تم استخدامها من قبل في الانحدار (1.3.9) تم الحصول على النتائج التالية :

$$\hat{Y}_i = -0.2610 - 2.3606D_{2i} - 1.7327D_{3i} + 0.8028X_i \qquad (4.6.9)$$

$$t = (-0.2357)^{**} \quad (-5.4873)^{*} \quad (-2.1803)^{*} \quad (9.9094)^{*}$$

$$R^2 = 0.2032 \qquad n = 528$$

حيث * تعني قيم الـ p-value أقل من 5% و** تعني قيم p-value أكبر من 5% يمكن للقارئ أن يتحقق من أن معاملات الجزء الثابت التفاضلية مفتوحة إحصائياً ولها الإشارة المتوقعة (لماذا؟). وإن التعليم له تأثير قوي طردي على الأجر بالساعة كما هو متوقع.

ومن (4.6.9) نرى أن متوسط الأجر المكتسب بالساعة للإناث أقل بحوالي 2,36$ وأيضاً متوسط الأجر المكتسب للعاملين ذوي البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية أقل أيضاً بحوالي 1,73$. دعنا الآن نعتبر نتائج النموذج (2.6.9) والتي تحتوي على الأثر التفاعلي الوهمي.

$$\hat{Y}_i = -0.26100 - 2.3606D_{2i} - 1.7327D_{3i} + 2.1289D_{2i}D_{3i} + 0.8028X_i$$
$$t = (-0.2357)^{**} \quad (-5.4873)^{*} \quad (-2.1803)^{*} \quad (1.7420)^{**} \quad (9.9095)^{**} \quad (5.6.9)$$
$$R^2 = 0.2032 \qquad n = 528$$

حيث : * تعني قيم p-value أقل من 5% و** تعني قيم الـ p-value أكبر من 5% والقارئ يمكن أن يلاحظ أن المتغيرين الوهميين المضافين مازال لهما معنوية إحصائية، ولكن المتغير الوهمي التفاعلي غير معنوي عند مستوى المعنوية 5%، قيمة الـ p-value الفعلية للجزء التفاعلي الوهمي حوالي 8% . اذا اعتبرت أن ذلك يعتبر احتمالاً قليلاً بشكل كاف فإن نتائج (5.6.9) يمكن تفسيرها كالتالي : بافتراض ثبات مستوى التعليم، إذا أضفنا المعاملات الوهمية الثلاثة سنحصل على : (2.1289+1.7327−2.3605−=)1.964− والذي يعني أن متوسط الأجر بالساعة للإناث ذوات البشرة غير البيضاء وغير المنتمين لأمريكا اللاتينية أقل بحوالي 1.964، وهذه القيمة بين 2.3605− (الفرق العائد إلى النوع فقط) و1.7327− (الفرق العائد إلى العرق فقط)

المثال السابق ، يوضح بشكل كبير، الدور الذي يلعبه التفاعل بين المتغيرات الوهمية، عندما يكون في النموذج اثنان أو أكثر من المتغيرات المنحدر عليها عبارة عن متغيرات نوعية.

من المهم ملاحظة أنه في النموذج (5.6.9) نفترض أن معدل الزيادة في الأجر المكتسب بالساعة بالنسبة للتعليم (بحوالي 80 سنتًا لكل سنة إضافية من المدرسة) يظل ثابتًا لكل طبقات النوع والعرق المختلفة. ولكن ذلك قد لا يكون متحققًا علمياً. إذا أردنا اختبار ذلك، لابد من إدخال معاملات الميل التفاضلية للنموذج (انظر تمرين 25.9).

7.9 استخدام المتغيرات الوهمية في التحليل الموسمي:
THE USE OF DUMMY VARIABLES IN SEASONAL ANALYSIS

العديد من السلاسل الزمنية الاقتصادية الموجودة بياناتها على أساس شهري أو ربع سنوي يظهر فيها نمط تغير موسمى (تحركات تذبذبية عادية). من الأمثلة على ذلك: المبيعات في محال عديدة عند أعياد الكريسماس، أو في أي أوقات إجازة مهمة، الطلب على النقود (أو النقود السائلة) من أرباب الأسر في أوقات الإجازات، الطلب على الآيس والمشروبات الباردة في فصل الصيف، أسعار المحاصيل بعد موسم جمعها مباشرة، الطلب على تذاكر الطيران وإلى ما غير ذلك من أمثلة عديدة.

عادة يكون من المحبب التخلص من هذا العامل الموسمي أو المكون الموسمي من السلسلة الزمنية، حيث يمكن للفرد دراسة المكونات الأخرى مثل الاتجاه العام[12].

العملية التي نقوم فيها بتخليص السلسلة من الأثر الموسمي من السلة الزمنية تسمى deseasonalization أو التعديل الموسمي والسلسلة الزمنية التي نحصل عليها في ذلك الوقت تسمى سلسلة زمنية مخلصة من الأثر الموسمي.

العديد من السلاسل الاقتصادية المهمة مثل معدل البطالة، ومؤشر سعر المستهلك (CPI)، ومؤشر سعر المنتج (PPT)، ومؤشر الإنتاج الصناعي منشورة في صورة سلاسل مخلصة من الأثر الموسمي.

هناك العديد من الطرق المستخدمة لتخليص السلسلة من الأثر الموسمي، ولكن سنتطرق فقط إلى واحدة فقط من هذه الطرق، وهي بالتحديد طريقة المتغيرات الوهمية[13].

لنوضح كيفية استخدام المتغيرات الوهمية لتخليص السلسلة الزمنية الاقتصادية من الأثر الموسمى، دعنا نستخدم البيانات الموجودة في جدول (3.9). هذا الجدول يعطي بيانات ربع سنوية من السنوات 1978 إلى 1995 والخاصة بمبيعات أربع سلع معمرة أساسية وهي غسالة الأطباق، مفرمة البواقي والفضل، الثلاجات وغسالات الملابس، كل البيانات معطاة بالآلاف. الجدول أيضاً يحتوي على بيانات الإنفاق على بعض السلع المعمرة الأخرى في 1982 بالبليون دولار.

لنوضح أسلوب المتغيرات الوهمية، دعنا نعتبر مبيعات الثلاجات خلال فترة الدراسة السابق ذكرها. ولكن دعنا أولاً ننظر إلى البيانات الموضحة بيانياً في الشكل (4.9). هذا الشكل يجعلنا نفترض وجود نمط موسمي في البيانات المعطاة بشكل ربع سنوي. لنرى حقيقة ذلك دعنا نعتبر النموذج التالي:

---

(12) السلسلة الزمنية قد تحتوي على أربعة مكونات: الموسمي، التكراري، الاتجاه العام، ومكون عشوائي بحت.

(13) للطرق المختلفة المستخدمة في التخلص من الأثر الموسمي، انظر على سبيل المثال في:
Francis X. Diebod, Elements of Forecasting, 2d ed., South-Western Publishers, 2001, Chap. 5,

$$Y_1 = \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_{3t} D_{3t} + \alpha_4 D_{4t} + u_t \qquad (1.7.9)$$

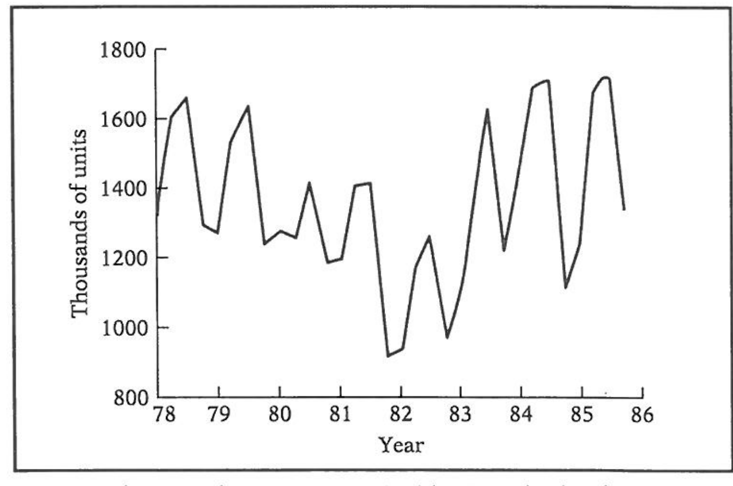
حيث: Y_1 = مبيعات الثلاجات (بالآلاف)، و D's هي المتغيرات الوهمية، وتأخذ 1 في حالة الربع السنوي محل الاهتمام، و0 بخلاف ذلك.

جدول (3.9) بيانات ربع سنوية عن مبيعات بعض الأجهزة الكهربائية (بالآلاف) والنفقات المصروفة على السلع المعمرة (I–1978 إلى IV–1985).

DISH	DISP	FRIG	WASH	DUR	DISH	DISP	FRIG	WASH	DUR
841	798	1317	1271	252.6	480	706	943	1036	247.7
957	837	1615	1295	272.4	530	582	1175	1019	249.1
999	821	1662	1313	270.9	557	659	1269	1047	251.8
960	858	1295	1150	273.9	602	837	973	918	262
894	837	1271	1289	268.9	658	867	1102	1137	263.3
851	838	1555	1245	262.9	749	860	1344	1167	280
863	832	1639	1270	270.9	827	918	1641	1230	288.5
878	818	1238	1103	263.4	858	1017	1225	1081	300.5
792	868	1277	1273	260.6	808	1063	1429	1326	312.6
589	623	1258	1031	231.9	840	955	1699	1228	322.5
657	662	1417	1143	242.7	893	973	1749	1297	324.3
699	822	1185	1101	248.6	950	1096	1117	1198	333.1
675	871	1196	1181	258.7	838	1086	1242	1292	344.8
652	791	1410	1116	248.4	884	990	1684	1342	350.3
628	759	1417	1190	255.5	905	1028	1764	1323	369.1
529	734	919	1125	240.4	909	1003	1328	1274	356.4

لاحظ أن Dish = غسالات الأطباق، Disp = مفرمة الفضلات والبواقي، FRTY = الثلاجات، Wash = غسالات الملابس، DUR = نفقات السلع المعمرة، بالبليون دولار في 1992.

المصدر: Business Statistics and Survey of Current Business, Department of Commerce (various issues).



شكل (4.9) مبيعات الثلاجات 1978 – 1985 (ربع سنوي)

لاحظ أنه حتى لا نقع في مصيدة المتغيرات الوهمية، نحدد متغيراً وهمياً لكل ربع من السنة ، مع حذف الجزء الثابت المقطوع من المحور الصادي . إذا كان هناك أثر موسمي لأي ربع من السنة، ستكون قيمة t الخاصة بعامل ا لمتغير الوهمي لهذا الربع معنوية إحصائياً[14] .

لاحظ أنه فى (1.7.9) نقوم فعلياً بعمل انحدار لـ Y على الجزء الثابت المقطوع من المحور الصادي، ولكننا نسمح بأجزاء ثابتة مختلفة لكل موسم (أي لكل ربع سنة) . كنتيجة لذلك المعامل الوهمي لكل ربع يعبر عن متوسط مبيعات الثلاجات في كل ربع أو موسم (لماذا؟) .

مثال 6.9

الموسمية في مبيعات الثلاجات : SEASONALITY IN REFRIGERATOR SALES

من البيانات الخاصة بمبيعات الثلاجات في جدول (3.9)، حصلنا على نتائج الانحدار التالية :

$$\hat{Y}_t = 1222.125D_{1t} + 1467.500D_{2t} + 1569.750D_{3t} + 1160.000D_{4t} \qquad (2.7.9)$$
$$t = \quad (20.3720) \qquad (24.4622) \qquad (26.1666) \qquad (19.3364)$$
$$R^2 = 0.5317$$

لاحظ أن : لا توجد قيم للأخطاء المعيارية للمعاملات المقدرة، فكل خطأ معياري يساوي 59.9904، حيث إن كل المتغيرات الوهمية تأخذ القيمة 1 أو 0 فقط .

معاملات α المقدرة في (2.7.9) تمثل متوسط مبيعات الثلاجات (بالآلاف) في كل موسم (أي ربع سنة) . وبالتالي متوسط مبيعات الثلاجات في الربع الأول حوالي 1222 (مقدرة بالآلاف)، وفي الربع الثاني 1468 وفي الربع الثالث 1570 والربع الأخير حوالي 1160 .

(14) هنا توجد ملاحظة فنية خاصة بالطريقة المستخدمة للمتغير الوهمي ، فقد تم تخصيص متغير وهمي لكل ربع سنة ، مع افتراض أنه في حالة وجود أثر موسمي سيكون ثابتًا وغير عشوائي . سنعود مرة أخرى إلى هذه النقطة في الجزء 7 من الكتاب لمناقشة السلاسل الزمنية الخاصة بالاقتصاد القياسي .

جدول (4.9) مبيعات الثلاجات في الولايات المتحدة (بالآلاف) ، 1978 - 1995 (ربع سنوي)

U.S. REFRIGERATOR SALES (THOUSANDS), 1978 – 1995 (QUARTERLY)

FRIG	DUR	D_2	D_3	D_4	FRIG	DUR	D_2	D_3	D_4
1317	252.6	0	0	0	943	247.7	0	0	0
1615	272.4	1	0	0	1175	249.1	1	0	0
1662	270.9	0	1	0	1269	251.8	0	1	0
1295	273.9	0	0	1	973	262.0	0	0	1
1271	268.9	0	0	0	1102	263.3	0	0	0
1555	262.9	1	0	0	1344	280.0	1	0	0
1639	270.9	0	1	0	1641	288.5	0	1	0
1238	263.4	0	0	1	1225	300.5	0	0	1
1277	260.6	0	0	0	1429	312.6	0	0	0
1258	231.9	1	0	0	1699	322.5	1	0	0
1417	242.7	0	1	0	1749	324.3	0	1	0
1185	248.6	0	0	1	1117	333.1	0	0	1
1196	258.7	0	0	0	1242	344.8	0	0	0
1410	248.4	1	0	0	1684	350.3	1	0	0
1417	255.5	0	1	0	1764	369.1	0	1	0
919	240.4	0	0	1	1328	356.4	0	0	1

لاحظ أن : FRIG : مبيعات الثلاجات بالآلاف .

DUR : نفقات السلع المعمرة ، بالبليون دولار في 1992 .

D_2 : 1 في الربع الثاني ، 0 بخلاف ذلك .

D_3 : 1 في الربع الثالث ، 0 بخلاف ذلك .

D_4 : 1 في الربع الرابع ، 0 بخلاف ذلك .

المصدر : Busines statistics and Survey of Current Business, Department of Commerce (various issues).

وبشكل بديل ، ممكن بدلاً من تحديد متغير وهمي لكل ربع ، وتجاهل الجزء الثالث حتى لا تقع في مصيدة المتغيرات الوهمية ، ممكن أن نحدد ثلاثة متغيرات وهمية فقط بالإضافة إلى جزء ثابت مقطوع من المحور الصادي . افترض أننا سنعامل الربع الأول على أنه الربع المرجعي ، ونحدد متغيرات وهمية للربع الثاني ، الثالث والرابع ، إذا قمنا بذلك سنحصل على نتائج الانحدار التالية (انظر جدول 4.9 لهيكل تجهيز البيانات) :

$$\hat{Y}_t = 1222.1250 + 245.3750D_{2t} + 347.6250D_{3t} - 62.1250D_{4t}$$

$$t = (20.3720)^* \quad (2.8922)^* \quad (4.0974)^* \quad (-0.7322)^{**} \qquad (3.7.9)$$

$$R^2 = 0.5318$$

حيث : * تعني قيم الـ p-value أقل من 5% ، ** تعني أن قيم p-value أكبر من 5%

حيث إننا نعامل الربع الأول على أنه الطبقة المرجعية ، المعاملات الملتحقة بالمتغيرات الوهمية المختلفة تسمى أجزاء ثابتة تفاضلية ، تمثل كمية الاختلاف في متوسط Y في الربع الذي يأخذ فيه المتغير الوهمي 1 عن نظيرها في الربع المرجعي .

وبشكل آخر إذا وضعنا معاملات المتغيرات الوهمية الموسمية سنحصل على الزيادة أو الانخفاض الموسمي في متوسط قيمة Y بالنسبة للموسم المرجعي. إذا جمعنا قيم الأجزاء الثابتة المختلفة التفاضلية إلى القيم المراجعية المساوية لـ 1222.125 سنحصل على متوسط العينة لكل أرباع السنة المختلفة. بهذه الطريقة سنحصل بالضبط على المعادلة (2.7.9) مع مراعاة الخطأ التقريبي.

دعنا الآن نستعرض القيم المختلفة في حالة التعامل مع ربع واحد كربع سنة مرجعي (3.7.9) يتضح أن متوسط قيمة Y للربع الرابع لا يختلف معنوياً عن متوسط قيمة الربع الأول، حيث إن المعامل الوهمي للربع الرابع غير معنوي إحصائياً. بالطبع هذا الاستنتاج سيتغير مع تغير الربع السنوي الذي يتم التعامل معه كربع مرجعي، ولكن الاستنتاج العام لن يتغير.

كيف يمكننا الحصول على السلسلة الزمنية لمبيعات الثلاجات مخلصة من الأثر الموسمي؟ يمكن القيام بذلك بسهولة. قم بتقدير قيمة الـY من النموذج (2.7.9) [أو (3.7.9)] لكل مفردة واطرح هذه القيم من القيم العقلية للـY، أي أننا نحصل على $(Y_t - \hat{Y}_t)$ والتي تمثل ببساطة بواقي الانحدار (2.7.9). سنستعرض ذلك في الجدول (5.9). [15]

من الذي تمثله هذه البواقي؟ تمثل المكونات الباقية للسلسلة الزمنية الخاصة بالثلاجات، أي بالتحديد، الاتجاه العام والأثر التكراري ومكون عشوائي (انظر إلى الملحوظة المذكورة في الهامش 15).

النموذجان (2.7.9) و(3.7.9) لا يشملان أي متغيرات مساعدة، هل ستتغير الصورة إذا استخدمنا متغيراً منحدراً عليه كمي في النموذج؟ حيث إن النفقات على السلع المعمرة لها دور كبير على الطلب على الثلاجات، دعنا نضيفهما إلى النموذج (3.7.9). بيانات نفقات السلع المعمرة بالبليون دولار في 1982 معطاة بالفعل في جدول (3.9). وسيمثل ذلك متغيرتي الكمي (X) في النموذج. نتائج الانحدار تصبح كالتالي:

(15) بالطبع يفترض ذلك أن أسلوب المتغيرات الوهمية وسيلة مناسبة لتخليص السلسلة الزمنية من الأثر الموسمي وأن السلسلة الزمنية (TS) يمكن التعبير عنها في صورة: $TS = s + c + t + u$ حيث s تمثل الأثر الموسمي، t تمثل الاتجاه العام، c الأثر التكراري وu مكون عشوائي. عموماً إذا كانت السلسلة الزمنية تأخذ الشكل، $TS = (s)(c)(t)(u)$ بمعنى أن المكونات الأربعة مضروبة في بعضها البعض يكون الأسلوب السابق غير مناسب لتخليص السلسلة الزمنية من الأثر الموسمي، وبالتالي فالطريقة السابقة تفترض أن المكونات الأربعة للسلسلة الزمنية مضافة (مجموعة) على بعضها البعض. ولكن هناك تفاصيل أكثر سنستعرضها عند مناقشة هذا الموضوع في الفصول الخاصة بالسلاسل الزمنية في الاقتصاد القياسي.

$$\hat{Y}_t = 456.2440 + 242.4976D_{2t} + 325.2643D_{3t} - 86.0804D_{4t} + 2.7734X_t$$
$$t = \quad (2.5593)^* \quad (3.6951)^* \quad (4.9421)^* \quad (-1.3073)^{**} \quad (4.4496)^* \qquad (4.7.9)$$
$$R^2 = 0.7298$$

حيث : * تعني أن قيم الـ p-values أقل من 5% و ** تعني أن الـ p-values أكبر من 5%.

جدول (5.9) انحدار مبيعات الثلاجات ، القيم الحقيقية والمقدرة وقيم البواقي [المعادلة 3.7.9]

REFRIGERATOR SALES REGRESSION ACTUAL, FITTED, AND RESIDUAL VALUES [EQ. (3.7.9)]

	Actual	Fitted	Residuals
1978-I	1317	1222.12	94.875
1978-II	1615	1467.50	147.500
1978-III	1662	1569.75	92.250
1978-IV	1295	1160.00	135.000
1979-I	1271	1222.12	48.875
1979-II	1555	1467.50	87.500
1979-III	1639	1569.75	69.250
1979-IV	1238	1160.00	78.000
1980-I	1277	1222.12	54.875
1980-II	1258	1467.50	−209.500
1980-III	1417	1569.75	−152.750
1980-IV	1185	1160.00	25.000
1981-I	1196	1222.12	−26.125
1981-II	1410	1467.50	−57.500
1981-III	1417	1569.75	−152.750
1981-IV	919	1160.00	−241.000
1982-I	943	1222.12	−279.125
1982-II	1175	1467.50	−292.500
1982-III	1269	1569.75	−300.750
1982-IV	973	1160.00	−187.000
1983-I	1102	1222.12	−120.125
1983-II	1344	1467.50	−123.500
1983-III	1641	1569.75	71.250
1983-IV	1225	1160.00	65.000
1984-I	1429	1222.12	206.875
1984-II	1699	1467.50	231.500
1984-III	1749	1569.75	179.250
1984-IV	1117	1160.00	−43.000
1985-I	1242	1222.12	19.875
1985-II	1684	1467.50	216.500
1985-III	1764	1569.75	194.250
1985-IV	1328	1160.00	168.000

Residual graph: − 0 +

والآن ضع في الاعتبار أننا نعامل الربع الأول على أنه الربع المرجعي. كما في (3.7.9)، نجد أن معاملات الجزء الثابت التفاضلية للربعين الثاني والثالث مختلفة بشكل

معنوي إحصائياً عن الربع الأول، أما الجزء الثابت الخاص بالربع الرابع والربع الأول منهما غير مختلفين إحصائياً. معامل X (نفقات السلع المعمرة) حوالي 2.77 مما يعني أنه مع وضع الأثر الموسمي في الاعتبار، إذا زادت النفقات على السلع المعمرة بدولار واحد فإنه في المتوسط تزداد مبيعات الثلاجات بحوالي 2,77 وحدة تقريباً 3 وحدات، ضع في الاعتبار أن الثلاجات مقاسة بالألف من الوحدة و X مقاسة في (1982) ببليون دولار.

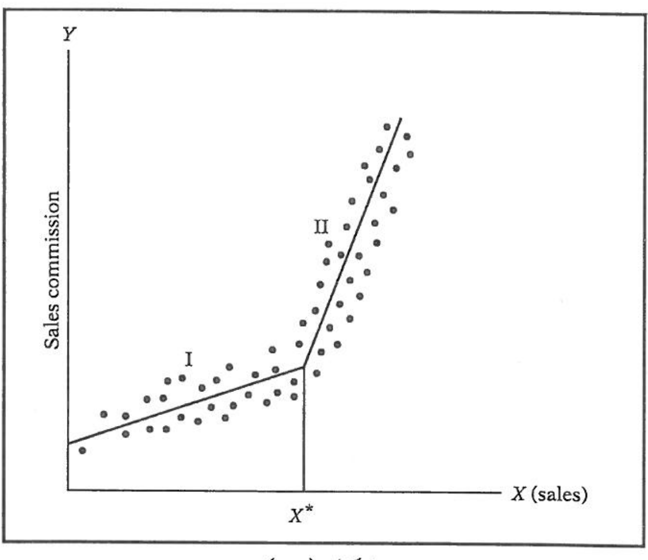
وهنا دعنا نطرح سؤالاً شيقاً: هل كون مبيعات الثلاجات يوجد فيها نمط موسمي، فإن ذلك يعني أن النفقات على السلع المعمرة لن يكون لها أيضاً نمط موسمي؟ كيف نضع في الاعتبار الموسمية التي قد توجد في X؟ الشيء المهم والخاص بـ (4.7.9) أن المتغير الوهمي في النموذج لا يخلص Y من الأثر الموسمي فقط بل يخلصنا أيضاً من الموسمية في X إذا وجدت. (هذا مثبت من خلال نظرية معروفة باسم Frisch-Waugh)[16] وبالتالي فإنه يمكننا القول بأننا قضينا على (الموسمية) في عصفوري (السلسلتين) بطوبة واحدة (أسلوب المتغير الوهمي).

إذن كيف تريد إثباتًا غير تقليدي للعبارة السابقة، اتبع الخطوات التالية: (1) قم بعمل انحدار لـ Y على المتغيرات الوهمية كما في (2.7.9) أو (3.7.9) واحتفظ بالبواقي، دعنا نطلق عليه S_1، هذه البواقي تمثل Y المخلصة من الأثر الموسمي، (2) قم بعمل انحدار مماثل لـ X واحصل على البواقي من هذا الانحدار واطلق عليها S_2، هذه البواقي تمثل X المخلصة من الأثر الموسمي، (3) اعمل انحداراً لـ S_1 على S_2 ستجد أن معامل الميل في هذا الانحدار هو نفسه معامل X في انحدار (4.7.9).

8.9 الانحدار الخطي الجزيئي: PIECEWISE LINEAR REGRESSION

لاستعراض استخدام آخر للمتغيرات الوهمية، دعنا نعتبر الشكل (5.9) والذي يعتبر مثالاً افتراضيًا لشركة تكافئ موظفي المبيعات لديها. فهذه الشركة تدفع عمولة بناء على المبيعات، على أساس أنه عندما يحقق الموظف مستوى أو حداً معينًا للمبيعات X^* أو أكثر منه، تكون له عمولة على هذا المستوى من المبيعات، أما البيع بأقل من هذا المستوى فلا توجد عمولة له (ملحوظة: ضع في الاعتبار أنه بالإضافة إلى المبيعات هناك عوامل أخرى تؤثر في عمولة المبيعات. دعنا نفترض أن هذه العوامل كلها ممثلة في حد الخطأ العشوائي). وبشكل أكثر تحديداً، فإن قيم الافتراض بأن عمولة المبيعات تتزايد خطياً مع المبيعات حتى المستوى المعين X^* فبعد هذا المستوى تتزايد أيضاً خطياً مع المبيعات ولكن بمعدل أكبر.

(16) للإثبات النظري، انظر Adrian C. Darnell, A Dictionary of Econometrics, Edward Elgar, Lyme, U.K., 1995, pp. 150–152.



شكل (5.9)

وبالتالي، فإن لدينا انحداراً خطيًا جزيئيًا يتكون من جزيئين خطيين واللذين نرمز لهما بـ I و II كما في الشكل (5.9) ودالة العمولة يتغير الميل الخاص بها بعد المستوى المحدد X^*. وبالتالي بمعلومية بيانات العمولة والمبيعات والمستوى المحدد X^*، يمكن استخدام أسلوب المتغيرات الوهمية لتقدير (الفرق) في الميل بين المرحلتين أو الجزيئين الخاصين بهذا الانحدار الخطي الموجود في الشكل (5.9). وهذا الانحدار يتم كالتالي:

$$Y_i = \alpha_1 + \beta_1 X_i + \beta_2 (X_i - X^*) D_i + u_i \qquad \textbf{(1.8.9)}$$

حيث: Y_i = عمولة المبيعات.

X_i = حجم المبيعات المسئول عنها موظف المبيعات.

X_t = المستوى المبدئي من المبيعات و المعروف أيضاً باسم العقدة (ويكون معلومًا مسبقاً)[17].

$D = 1$ إذا كانت $X_i > X^*$

$= 0$ إذا كانت $X_i < X^*$

(17) القيمة المبدئية قد لا تكون دائماً معروفة. عموماً أسلوب An ad hoc يعتمد على رسم المتغير التابع مع المتغيرات المفسرة وملاحظة ما إذا كان هناك تغير جاد في العلاقة بعد قيمة معينة لـ X (أي X^*). أسلوب تحليلي لمعرفة نقطة التغير معروف باسم نماذج الانحدار المتقلبة. لكن هذا الموضوع يعتبر موضوعاً متقدماً ويمكن القراءة في المزيد من تفاصيله في

Thomas Fomby, R. Carter Hill, and Stanley Johnson, Advanced Econometric Methods, Springer-Verlag, New York, 1984, Chap. 14.

مع افتراض أن $E(u_i) = 0$، والآن يمكن مباشرة ملاحظة التالي:

$$E(Y_i \mid D_i = 0, X_i, X^*) = \alpha_1 + \beta_1 X_i \qquad (2.8.9)$$

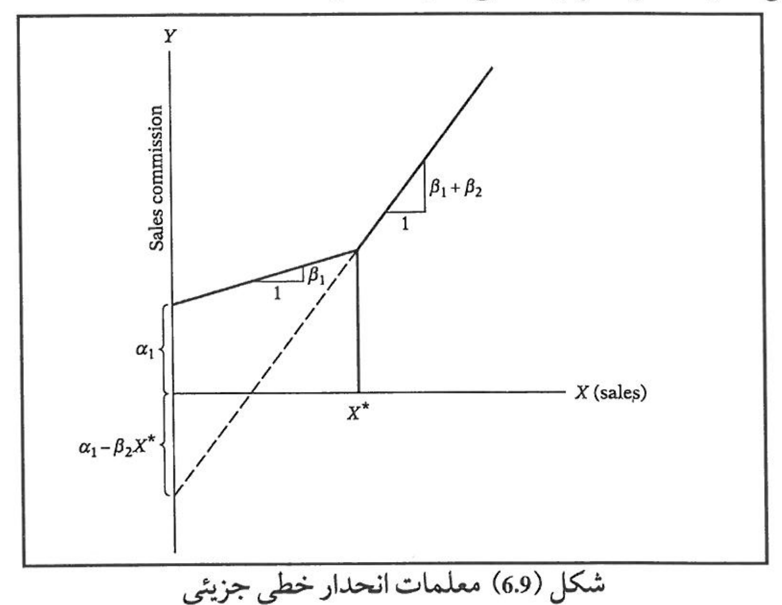
والذي يمثل أيضاً متوسط عمولات المبيعات حتى المستوى المطلوب X^*

و
$$E(Y_i \mid D_i = 1, X_i, X^*) = \alpha_1 + \beta_2 X^* + (\beta_1 + \beta_2) X_i \qquad (3.8.9)$$

يمثل متوسط عمولات المبيعات بعد المستوى المطلوب X^*.

وبالتالي، فإن β_1 تمثل ميل خط الانحدار في المرحلة I، و $\beta_1 + \beta_2$ تمثل ميل خط الانحدار في المرحلة II من الانحدار الخطي الجزيئي الموضح في الشكل (5.9). لاختبار الفرض القائل بأنه لا يوجد تغيير في الانحدار عند القيمة المبدئية X^* يمكن إجراؤه بسهولة بملاحظة المعنوية الإحصائية لمعامل الميل المقدر التفاضلي $\hat{\beta}_2$ (انظر الشكل 6.9).

وبمحض المصادفة يعتبر الانحدار الخطي الجزيئي الذي ناقشناه الآن يعتبر مثالاً لفئة من الدوال أكثر عمومية تسمى الدوال Spline [18].



شكل (6.9) معلمات انحدار خطي جزيئي

(18) لمزيد من التفاصيل عن Splines (أي متعددة حدود جزئية من الدرجة k) انظر في
Douglas C. Montgomery, Elizabeth A. Peck, and G. Geoffrey Vining, Introduction to Linear Regression Analysis, John Wiley & Sons, 3d ed., New York, 2001, pp. 228–230.

مثال 7.9

التكلفة الكلية وعلاقتها بالناتج : TOTAL COST IN RELATION TO OUTPUT

كمثال تطبيقي للانحدار الخطي الجزيئي، دعنا نتعامل مع بيانات افتراضية عن التكلفة الكلية - الناتج الكلي الموجود في الجدول (6.9). ودعنا نفترض أننا نعلم أن الميل يتغير عند مستوى 5500 وحدة.

دع Y الموجودة في (4.8.9) تمثل التكلفة الكلية، وX الناتج الكلي، سنحصل على النتائج التالية:

$$\hat{Y}_i = -145.72 \quad + \; 0.2791X_i + \; 0.0945(X_i - X_i^*)D_i$$
$$t = \quad (-0.8245) \quad (6.0669) \quad (1.1447) \qquad (4.8.9)$$
$$R^2 = 0.9737 \qquad X^* = 5500$$

كما يتضح من هذه النتائج، التكلفة الحدية للإنتاج حوالي 28 سنتًا للوحدة، وعلى الرغم من أن حوالي 37 سنتًا (28 + 9) للناتج الزائد عن 5500 وحدة، إلا أن الفرق بين القيمتين ليس معنويًا إحصائياً، حيث إن المتغير الوهمي غير معنوي عند مستوى المعنوية 5%. ولكن عمليين يمكن عمل انحدار للتكلفة الكلية على الناتج الكلي مع حذف المتغير الوهمي.

جدول (6.9) مثال تطبيقي للانحدار الخطي الجزيئي

Total cost, dollars	Output, units
256	1,000
414	2,000
634	3,000
778	4,000
1,003	5,000
1,839	6,000
2,081	7,000
2,423	8,000
2,734	9,000
2,914	10,000

9.9 نماذج انحدار البيانات الجدولية: PANEL DATA REGRESSION MODELS

تذكر أنه في الفصل (1)، ناقشنا أنواع البيانات المختلفة والمتاحة في التحليل التطبيقي، مثل البيانات المقطعية، بيانات السلاسل الزمنية، والبيانات التجميعية (خليط من بيانات السلاسل الزمنية والبيانات المقطعية والبيانات الجدولية). أسلوب المتغيرات الوهمية يمكن استخدامه بسهولة في البيانات التجميعية والجدولية. وحيث إن هناك تزايدًا في استخدام البيانات الجدولية في المجال التطبيقي، سنتطرق إلى هذا الموضوع بالتفصيل في الفصل (16).

10.9 بعض الجوانب الفنية في أسلوب المتغير الوهمي : SOME TECHNICAL ASPECTS OF THE DUMMY VARIABLE TECHNIQUE

تفسيرات المتغيرات الوهمية في الانحدارات شبه اللوغاريتمية في الفصل (6)، قمنا بمناقشة نماذج log-lin حيث يأخذ المتغير المنحدر الشكل اللوغاريتمي والمتغيرات المنحدر عليها تكون خطية. في شكل هذه النماذج، معاملات الميل للمتغيرات المنحدر عليها تعبر عن مرونة نسبية، بمعنى أنها تمثل نسبة التغير في المتغير المنحدر لكل تغير بمقدار الوحدة في المتغير المنحدر عليه. هذا التعريف يكون سليمًا في حالة ما إذا كان المتغير المنحدر عليه متغير كمي فقط. ماذا سيحدث إذا كان المتغير المنحدر عليه متغير وهمي؟ للتحديد دعنا نستعرض النموذج التالي:

$$\ln Y_i = \beta_1 + \beta_2 D_i + u_i \qquad (1.10.9)$$

حيث: Y = معدل الأجر بالساعة ($)، $D = 1$ إذا كان العامل أنثى، و 0 إذا كان ذكرًا.

كيف يمكننا تفسير مثل هذا النموذج؟ افترض أن $E(u_i) = 0$، سنحصل على:

دالة الأجر للعاملين الذكور :

$$E(\ln Y_i / D_i = 0) = \beta_1 \qquad (2.10.9)$$

دالة الأجر للعاملات الإناث :

$$E(\ln Y_i / D_i = 1) = \beta_1 + \beta_2 \qquad (3.10.9)$$

وبالتالي، فإن الجزء المقطوع من المحور الصادي β_1، يمثل متوسط لوغاريتم الأجر بالساعة، ومعامل «الميل» يمثل الفرق بين متوسط لوغاريتم الأجر بالساعة بين الذكور والإناث. ويتضح ضعف التفسيرات السابقة، ولكن إذا استخدمنا الدالة اللوغاريتمية العكسية لـ β_1، فإننا لن نحصل على متوسط الأجر بالساعة للذكور العاملين، ولكن سنحصل على وسيط الأجر. وكما تعلم الوسط والوسيط والمنوال تعتبر ثلاثة مقاييس مختلفة للشركة المركزية لأي متغير عشوائي. وإذا استخدمنا أيضاً الدالة اللوغاريتمية العكسية لـ $(\beta_1 + \beta_2)$ سنحصل على وسيط الأجر بالساعة للعاملات الإناث.

مثال 8.9

لوغاريتم الأجور بالساعة وعلاقته بالنوع :

LOGARITHM OF HOURLY WAGES IN RELATION TO GENDER

لتوضيح (1.10.9) دعنا نستخدم بيانات المثال 2.9. نتائج الانحدار باستخدام 528 مشاهدة هي كالتالي :

$$\widehat{\ln Y_i} = 2.1763 - 0.2437 D_i$$
$$t = (72.2943)^* \quad (-5.5048)^* \qquad (4.10.9)$$
$$R^2 = 0.0544$$

حيث : * تعني أن قيم *p*-valus عملياً تساوي الصفر.

باستخدام الدالة اللوغاريتمية العكسية للقيمة 2.1763 نحصل على ($)8.8136 وهذا يمثل وسيط الأجر بالساعة للعاملين الذكور ، وباستخدام الدالة اللوغاريتمية العكسية لـ [1.92857 = (2.1763–0.2437)] نحصل على ($)6.8796 وهذا يمثل وسيط الأجر بالساعة للعاملات الإناث ، وبالتالي وسيط أجر العاملات الإناث أقل بحوالي 21.94% مقارنةً بنظيرهن من الذكور [8.8136 / (8.8136–6.8796)] .

ومما يثير الاهتمام أيضاً أننا يمكننا الحصول على المرونة النسبية للمتغير المنحدر الوهمي مباشرةً بالطريقة المقترحة من كل من Halrorsong Palmquit [19] . فباستخدام الدالة اللوغاريتمية العكسية (للأساس *e*) للمعامل الوهمي المنحدر ويطرح 1 منها وبضرب الفرق في 100 (لتعليل ذلك، انظر الملحق 1.A.9) . وبالتالي إذا أخذنا اللوغاريتم العكسي لـ 5.2163– سنحصل على 0.78366 وبطرحه من 1 نحصل على 0.2163– وبضربه في 100 نحصل على 21.63–% مما يعني أن وسيط أجر العاملة الأنثى ($D = 1$) أقل من نظيرها الرجل بحوالي 21.63% . وهذه النتيجة حصلنا عليها من قبل مع تجاهل الخطأ التقريبي.

المتغيرات الوهمية واختلاف التباين: Dummy Variables and Heteroscedasticity

بالعودة إلى مثالنا الخاص بانحدار الدخل - الإدخار للولايات المتحدة خلال الفترة 1970 - 1981 و1982 - 1995 والفترة مجمعة 1995 - 1970 . لاختيار الاستقرار الهيكلي با ستخدام أسلوب المتغيرات الوهمية ، افترضنا أن الأخطاء لها تباينات متساوية في الفترتين، أي أن $\text{var}(u_{1i}) = \text{var}(u_{2i}) = \sigma^2$. وهذا الفرض لابد من تحققه أيضاً في اختبار Chow . إذا كان الفرض غير متحقق - أي أن تباينات الأخطاء في

(19) Robert Halvorsen and Raymond Palmquist, "The Interpretation of Dummy Variables in Semilogarithmic Equations, "American Economic Review, vol. 70, no. 3, pp. 474–475.

الفترتين غير متساوية - يكون من المحتمل الحصول على استنتاجات خاطئة. وبالتالي يجب أولاً أن يتم التحقق من تساوي التباين في الفترتين، باستخدام أساليب إحصائية مناسبة. وعلى الرغم من أننا سنناقش هذا الموضوع بالتفصيل في الفصل الخاص باختلاف التباين، إلا أننا في الفصل (8)، أوضحنا كيف يمكن استخدام اختبار F للتحقق من هذا الفرض.[20] (انظر إلى مناقشتنا لاختبار Chow وأسلوب المتغير الوهمي التي استعرضناها من قبل قد لا تكون سليمة. بالطبع هدفنا هنا هو توضيح الطرق المختلفة التي يمكن للفرد أن يعتمد عليها لعلاج أي مشكلة (مثل مشكلة الاستقرار الهيكلي). في أي تطبيق عملي، قد لا تكون هذه الأساليب سليمة. ولكن هذا قد يحدث لأي أساليب إحصائية بالطبع يمكن للفرد أن يتعامل بأي أسلوب مناسب ليعالج المشكلة التي تواجهه في التطبيق العملي، وسوف نقوم بذلك في الفصل الخاص باختلاف التباين لاحقاً (عموماً انظر تمرين 8.2.9).

المتغيرات الوهمية والارتباط الذاتي: Dummy Variables and Autocorrelation

بالإضافة إلى فرض ثبات التباين، فإن نموذج الانحدار الخطي التقليدي يفترض أن حد الخطأ في نماذج الانحدار غير مرتبط. ولكن ماذا سيحدث إذا لم يتحقق ذلك، خصوصاً في النماذج التي تشتمل على متغيرات منحدر عليها وهمية؟ وحيث إننا سنناقش بعمق موضوع الارتباط الذاتي في الفصل الخاص بالارتباط الذاتي أننا سنؤجل مناقشة هذا الموضوع حتى نصل إلى هذا الفصل.

ماذا سيحدث إذا كان المتغير التابع متغيراً وهمياً؟
What Happens if the dependent Variable is a Dummy Variable?

حتى الآن ناقشنا النماذج التي يكون فيها المتغير المنحدر متغيراً كميًا، والمتغيرات المنحدر عليها كمية أو نوعية أو كليهما معاً. ولكن هناك بعض المواقف التي يكون فيها المتغير المنحدر نفسه نوعي أو وهمي. فعلى سبيل المثال، القرار الخاص بالفرد ليكون ضمن القوى العاملة. قرار المشاركة هنا من نوع البيانات الخاصة بنعم أو لا. نعم إذا كان الفرد قرر المشاركة ولا بخلاف ذلك. وبالتالي متغير المشاركة في القوى

(20) خطوات اختبار Chow يمكن القيام بها حتى في حالة اختلاف التباين، ولكن لابد في هذه الحالة من استخدام اختبار Wald. الخطوات الرياضية لهذا الاختبار معقدة إلى حد ما. ولكن في الفصل الخاص باختلاف التباين سنستعرض هذه النقطة بالتفصيل.

العاملة هو متغير وهمي. بالطبع قرار المشاركة في القوى العاملة يعتمد على عوامل عديدة، مثل معدل الأجر عند بداية العمل، التعليم، وظروف سوق العمل (والمقاسة بمعدل البطالة).

هل مازال يمكننا استخدام الـ OLS لتقدير نماذج الانحدار عندما يكون المتغير المنحدر وهميًا؟ نعم نظرياً يمكننا القيام بذلك. ولكن هناك العديد من المشاكل الإحصائية التي قد يواجهها الفرد في مثل هذه النماذج. وبما أن هناك بدائل لتقدير OLS والتي لا تواجه مثل هذه المشاكل، فإننا سنستعرض هذا الموضوع في فصل لاحق (انظر الفصل 15 على نماذج logit و probit). في هذا الفصل، سنناقش أيضاً النماذج التي يكون المتغير المنحدر له أكثر من حالتين أو طبقتين، مثل قرار الذهاب إلى العمل بالسيارة، الأتوبيس أو القطار أو قرار العمل كل الوقت أو بعض الوقت أو عدم العمل على الإطلاق. مثل هذه النماذج تسمى نماذج المتغير التابع المتعدد الحدود مقارنةً مع نماذج المتغير التابع ثنائي الحدود، حيث يكون المتغير التابع له طبقتان اثنتان فقط.

11.9 مواضيع للدراسات المستقبلية: TOPICS FOR FURTHER STUDY

هناك العديد من المواضيع المرتبطة بالمتغيرات الوهمية، والتي تمت مناقشتها في الأدبيات، ولكنها متقدمة بعض الشيء مثل (1) نماذج المعالم المتغيرة أو العشوائية، (2) نماذج الانحدار المتقلبة، (3) النماذج غير المتزنة.

في نماذج الانحدار التي درستها، يفترض أن المعامل β's غير معروف، ولكنه ثابت. نماذج المعاملات العشوائية - وهناك العديد منها - تفترض أن β's عشوائية أيضاً. يمكنك الرجع لـ Cwany كمرجع رئيسي لهذا الموضوع.[21]

في نماذج المتغير الوهمي باستخدام كل من الميل، والأجزاء الثابتة التفاضلية، يفترض صراحةً أننا نعمل نقطة التغيير. وبالتالي في مثالنا الخاص بالدخل - الادخار للفترة 1970 - 1995.

قسمنا الفترة إلى 1970 - 1981 و1982 - 1995. الفترة السابقة واللاحقة لنقطة التغير موجودة على أساس أنه في العام 1982 تغيرت العلاقة بين الدخل والادخار.

---

(21) P. A. V. . Swamy, Statistical Inference in Random Coefficient Regression Models, Springer-Verlag, Berlin, 1971.

أحياناً يكون من الصعب تحديد هذه النقطة، ومعرفة بالضبط فترة أو نقطة التغيير. أسلوب نماذج الانحدار المتقلبة (SRM) تم تطويره حتى يستخدم في مثل هذه الحالة. SRM يتعامل مع نقطة التغيير كمتغير عشوائي، ومن خلال عملية تكرارية تتحدد نقطة التغيير. هذا الأسلوب يعود الفضل فيه إلى Quande و Gold Feld.[22]

أساليب تقدير خاصة نحتاج إليها للتعامل مع ما هو معروف باسم مواقف عدم الاتزان، وهذه المواقف التي يحدث فيها عدم اتزان أو وضوح للسوق (أي أن الطلب لا يساوي العرض) المثال التقليدي هو العرض والطلب الخاص بسلعة ما. الطلب على السلعة هو دالة في سعرها وعوامل أخرى، والعرض الخاص بالسلعة هو دالة في سعرها وعوامل أخرى، هذه العوامل قد تختلف عن الموجود في دالة الطلب. وبالتالي الكمية المعروضة والمطلوبة لهذه السلعة ليست بالضرورة متساوية مع التي نحصل عليها عند تساوي العرض مع الطلب، وذلك يؤدي إلى وجود حالة عدم الاتزان. لمناقشة أكثر تفصيلاً لنماذج عدم الاتزان، يمكن للقارئ أن يرجع إلى Quandt.[23]

12.9 الخلاصة والاستنتاجات : SUMMARY AND CONCLUSIONS

1 - المتغيرات الوهمية التي تأخذ القيم 1 و0 (أو تحويلتهما الخطية) تعني استخدام متغيرات منحدر عليها نوعية في نماذج الانحدار.

2 - المتغيرات الوهمية هي أداة لتقسيم البيانات، بمعنى أننا نقسم العينة إلى مجموعات جزيئية متعددة بناء على الاتجاهات المختلفة (النوع، الحالة الاجتماعية، العرق، الديانة، إلى ما غير ذلك) ويسمح بشكل صريح بإجراء انحدارات فردية لكل مجموعة جزيئية.

إذا كانت هناك فروق في ردود المتغير المنحدر بناء على التغير في المتغيرات النوعية في المجموعات الجزيئية المختلفة، فإن ذلك سيظهر في الفروق بين معاملات الأجزاء الثانية المقطوعة من المحور الصادي، ومعاملات الميل الخاص بانحدارات المجموعات الجزيئية المختلفة.

3- على الرغم من سهولة أسلوب المتغيرات الوهمية، إلا أنه يجب التعامل معه بحذر. أولاً إذا كان هناك جزء ثابت في الانحدار، فإن عدد المتغيرات الوهمية

(22) S. Goldfeld and R. Quandt, Nonlinear Methods in Econometrics, North Holland, Amsterdam, 1972.

(23) Richard E. Quandt, The Econometrics of Disequilibrium, Basil Blackwell, New York, 1988.

لابد أن يكون أقل بواحد من عدد طبقات المتغير النوعي. ثانياً، المعامل المرتبط بالمتغير الوهمي لابد أن يفسر باستمرار على أساس علاقته بالمجموعة المرجعية أو مجموعة الأساس - أي المجموعة التي يتم التعبير عنها بالقيمة صفر.

المجموعة أو الطبقة المرجعية يعتمد اختبارها على الهدف من البحث محل الدراسة ، وأخيراً إذا كان النموذج يحتوي على العديد من المتغيرات النوعية بطبقات عديدة، فإن استخدام المتغيرات الوهمية سيستهلك عدداً كبيراً من درجات الحرية. ولهذا فإنه يجب على الفرد دائماً أن يوازن بين عدد المتغيرات الوهمية اللازمة والعدد الكلي من المفردات المتاحة للتحليل الإحصائي.

4 - هذا الفصل بالإضافة إلى تضمنه العديد من التطبيقات، فإنه يشتمل أيضاً على التالي (1) مقارنة انحدارين أو أكثر (2) تخليص بيانات السلسلة الزمنية من الأثر الموسمي (3) التداخل بين المتغيرات الوهمية (4) تفسيرات المتغيرات الوهمية في النماذج شبه اللوغاريتمية و(4) نماذج الانحدار الخطي الجزيئية.

5 - أوضحنا أيضاً بعض النقاط المهمة والخاصة با لمتغيرات الوهمية، واختلاف التباين والارتباط الذاتي. ولكن بما أننا سنتطرق إلى هذين الموضوعين بالتفصيل في فصول قادمة، سنعاود استعراض هذه النقاط بالتفصيل في حينه.

تمارين : EXERCISES

أسئلة : Questions

1.9 إذا كانت لديك بيانات شهرية خلال عدد من السنوات، كم عدد المتغيرات الوهمية اللازمة لاختبار الفروض التالية:

(a) الشهور الـ12 من السنة يوجد فيها نمط موسمي.

(b) الشهور فبراير، أبريل، يونيو، أغسطس، أكتوبر، وديسمبر فقط يوجد فيها نمط موسمي.

2.9 اعتبر نتائج الانحدار التالية (نسب t معطاة بين الأقواس)(*):

(*) Jane Leuthold, "The Effect of Taxation on the Hours Worked by Married Women," Industrial and Labor Relations Review, no. 4, July 1978, pp. 520–526.

(تم تغيير الرموز حتى تتناسب مع شكل الكتاب الحالي)

$$\hat{Y}_i = 1286 + 104.97X_{2i} - 0.026X_{3i} + 1.20X_{4i} + 0.69X_{5i}$$
$$t = \quad (4.67) \quad (3.70) \quad (-3.80) \quad (0.24) \quad (0.08)$$
$$-19.47X_{6i} + 266.06X_{7i} - 118.64X_{8i} - 110.61X_{9i}$$
$$(-0.40) \quad (6.94) \quad (-3.04) \quad (-6.14)$$
$$R^2 = 0.383 \quad n = 1543$$

حيث : Y = عدد ساعات العمل السنوية للزوجة، محسوب على أساس عدد ساعات العمل العادية في السنة مضافًا إليه أسابيع البحث عن عمل .

X_2 = متوسط الكسب الحقيقي بالساعة بعد استقطاع الضرائب للزوجة .

X_3 = الكسب السنوي الحقيقي للزوج في السنة الماضية بعد استقطاع الضرائب .

X_4 = عمر الزوجة بالسنوات .

X_5 = عدد سنوات الدراسة الكاملة للزوجة .

X_6 = متغير يعبر عن رأي البحوث في مسألة ما، 1 = إذا كان المبحوث يرى أنه من حق المرأة أن تعمل إذا رغبت في ذلك ووافق زوجها، 0 = بخلاف ذلك .

X_7 = متغير يعبر عن رأي المبحوث في مسألة ما، 1 = إذا كان المبحوث الزوج يفضل عمل زوجته، 0 = بخلاف ذلك .

X_8 = عدد الأطفال الأقل من 6 سنوات .

X_9 = عدد الأطفال في المرحلة العمرية من 6 إلى 13 سنة .

(a) هل الإشارات الخاصة بمعاملات المتغيرات المنحدر عليها غير الوهمية لها أي معنى اقتصادي؟ علل إجابتك .

(b) كيف يمكنك تفسير المتغيرات الوهمية X_6 و X_7؟ هل هذه المتغيرات الوهمية معنوية إحصائياً؟ بما أن العينة كبيرة نسبياً، فإنه يمكنك استخدام القاعدة "2-t" للإجابة على هذا السؤال .

(c) في رأيك . . ما هي الأسباب التي تجعل متغيرات العمر والتعليم عوامل غير معنوية في قرار مشاركة المرأة في القوة العاملة والذي تناولته هذه الدراسة؟

3.9 اعتبر نتائج الانحدار التالية(*) (البيانات الفعلية موجودة في الجدول 7.9)

(*) Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958–1971," The Economic Journal, vol. 82, March 1972, pp. 195–202.

$$\widehat{\text{UN}}_t = 2.7491 + 1.1507D_t - 1.5294V_t - 0.8511(D_tV_t)$$
$$t = (26.896) \quad (3.6288) \quad (-12.5552) \quad (-1.9819)$$
$$R^2 = 0.9128$$

جدول (7.9) مصفوفة بيانات الانحدار الموجود في تمرين 3.9

Data Matix for Regnession, in exerase 9.3

Year and quarter	Unem-ployment rate UN, %	Job vacancy rate *V*, %	*D*	*DV*
1958-IV	1.915	0.510	0	0
1959-I	1.876	0.541	0	0
-II	1.842	0.541	0	0
-III	1.750	0.690	0	0
-IV	1.648	0.771	0	0
1960-I	1.450	0.836	0	0
-II	1.393	0.908	0	0
-III	1.322	0.968	0	0
-IV	1.260	0.998	0	0
1961-I	1.171	0.968	0	0
-II	1.182	0.964	0	0
-III	1.221	0.952	0	0
-IV	1.340	0.849	0	0
1962-I	1.411	0.748	0	0
-II	1.600	0.658	0	0
-III	1.780	0.562	0	0
-IV	1.941	0.510	0	0
1963-I	2.178	0.510	0	0
-II	2.067	0.544	0	0
-III	1.942	0.568	0	0
-IV	1.764	0.677	0	0
1964-I	1.532	0.794	0	0
-II	1.455	0.838	0	0
-III	1.409	0.885	0	0
-IV	1.296	0.978	0	0
1965-I	1.201	0.997	0	0
-II	1.192	1.035	0	0
-III	1.259	1.040	0	0
-IV	1.192	1.086	0	0
1966-I	1.089	1.101	0	0
-II	1.101	1.058	0	0
-III	1.243	0.987	0	0
-IV	1.623	0.819	1	0.819
1967-I	1.821	0.740	1	0.740
-II	1.990	0.661	1	0.661
-III	2.114	0.660	1	0.660
-IV	2.115	0.698	1	0.698
1968-I	2.150	0.695	1	0.695
-II	2.141	0.732	1	0.732
-III	2.167	0.749	1	0.749
-IV	2.107	0.800	1	0.800
1969-I	2.104	0.783	1	0.783
-II	2.056	0.800	1	0.800
-III	2.170	0.794	1	0.794
-IV	2.161	0.790	1	0.790
1970-I	2.225	0.757	1	0.757
-II	2.241	0.746	1	0.746
-III	2.366	0.739	1	0.739
-IV	2.324	0.707	1	0.707
1971-I	2.516*	0.583*	1	0.583*
-II	2.909*	o.524*	1	0.524*

* تقديرات أولية

المصدر : Damodar Gujarati, "The Behaviour of Unemployment and Unfilled Vacancies: Great Britain, 1958–1971," The Economic Journal, vol. 82, March 1972, p. 202.

حيث : UN = معدل البطالة ، %

V = معدل الوظائف الخالية ، %

D = 1 ، للفترة التي تبدأ من 1966–IV

= 0 ، للفترة قبل 1966–IV

t = الزمن ، مقاس بربع السنة .

لاحظ أن: في الربع الرابع من 1966، وزارة القوى العاملة حررت نظام التأمين القومي عن طريق تبديل نظام المعدل الثابت للتعويضات للبطالة قصيرة الأجل إلى نظام مختلط من المعدل الثابت والتعويضات المرتبطة بالدخل (السابق) مما يزيد من مستوى تعويضات البطالة.

(a) ما هي توقعاتك المسبقة عن العلاقة بين معدل البطالة ومعدل الوظائف الخالية؟

(b) بافتراض ثبات معدل الوظائف الخالية، ما هو متوسط معدل البطالة في الفترة التي تبدأ من الربع الرابع في 1966؟ هل هناك فرق إحصائي بين تلك الفترة والفترة السابقة للربع الرابع في 1966؟ علل إجابتك.

(c) هل الميل الخاص بالفترة قبل الربع الرابع في 1966 والميل الخاص بالفترة بعد الربع الرابع في 1966 مختلفين؟ علل إجابتك.

(b) هل من الممكن الاستنتاج من هذه الدراسة أن التعويضات المرتفعة للبطالة تؤدي إلى زيادة في معدلات البطالة؟ هل هناك منطق اقتصادي وراء ذلك؟

4.9 بناء على بيانات سنوية خلال الفترة 1972 إلى 1979 قام Willian Nordhaus بتقدير النموذج التالي لتفسير سلوك سعر البترول الخاص بمنظمة الأوبك (الأخطاء المعيارية معطاة بين الأقواس)(*)

$$\hat{y}_t = 0.3x_{1t} + 5.22x_{2t}$$
$$se = (0.03) \quad (0.50)$$

حيث: y = الفرق بين سعر العام الحالي والسابق (دولار لكل برميل).

x_1 = الفرق بين سعر السنة الحالية للبرميل وسعر الأويك في السنة السابقة.

x_2 = 1 لسنة 1974 و 0 بخلاف ذلك.

فسر هذه النتائج، ووضح النتائج بيانياً. ما الذي تقترحه هذه النتائج عن قوة الاحتكار الخاصة بالأويك؟

5.9 اعتبر النموذج التالي

$$Y_i = \alpha_1 + \alpha_2 D_i + \beta X_i + u_i$$

حيث: Y = الدخل السنوي للأستاذ الجامعي.

X = عدد سنوات الخبرة في مجال التدريس.

D = متغير وهمي خاص بالنوع.

(*) "Oil and Economic Performance in Industrial Countries," Brookings Papers on Economic Activity, 1980, pp. 341–388.

اعتبر الآن الثلاث طرق التالية المختلفة لتعريف المتغير الوهمي :

(a) $D = 1$ للذكر، 0 للأنثى.

(b) $D = 1$ للأنثى، 2 للذكر.

(c) $D = 1$ للأنثى، 1 = للذكر.

فسر نتائج الانحدار الخاصة بهذا النموذج وفقاً لكل طريقة تعريف للمتغير الوهمي. هل هناك طريقة تعريف مفضلة عن الأخرى؟ علل إجابتك.

6.9 بالرجوع إلى الانحدار (3.7.9)، كيف يمكنك اختبار الفرض القائل أن معاملات كل من D_2 و D_3 متساوية؟ ومعاملات D_2 وD_4 أيضاً متساوية؟ إذا كان معامل D_3 مختلفًا معنوياً عن D_2، ومعامل D_4 مختلفًا معنوياً أيضاً عن D_2، هل هذا يعني أن معاملات D_3 وD_4 مختلفة؟

ملحوظة مساعدة:

$$\text{var}(A \pm B) = \text{var}(A) + \text{var}(b) \pm 2\,\text{cov}(A, B)$$

7.9 بالرجوع إلى مثال الدخل - الادخار في الولايات المتحدة والذي ناقشناه في هذا الفصل

(a) كيف يمكنك الحصول على الأخطاء القياسية لمعاملات الانحدار المعطاة في (5.5.9) و(6.5.9)، والتي حصلنا عليها من الانحدار التجميعي (4.5.9)؟

(b) للحصول على إجابات عددية، ما هي المعلومات الإضافية التي قد تحتاج إليها؟

8.9 في دراسة عن تقييم واختبار بساعات العمل الخاصة بـ FDICC (المؤسسة الفيدرالية لإيداعات التأمين) على 91 مؤسسة بنكية، قام R.J.Miller بتقدير الدالة التالية :(*)

$$\widehat{\ln Y} = \underset{}{2.41} + \underset{(0.0477)}{0.3674 \ln X_1} + \underset{(0.0628)}{0.2217 \ln X_2} + \underset{(0.0287)}{0.0803 \ln X_3}$$

$$\underset{(0.2905)}{-0.1755 D_1} + \underset{(0.1044)}{0.2799 D_2} + \underset{(0.1657)}{0.5634 D_3} - \underset{(0.0787)}{0.2572 D_4}$$

$$R^2 = 0.766$$

(*) "Examination of Man-Hour Cost for Independent, Joint, and Divided Examination Programs," Journal of Bank Research, vol. 11, 1980, pp. 28–35.

حيث Y = محقق الـ FDIC والمختبر لساعات العمل.

X_1 = إجمالي أصول البنك.

X_2 = عدد المكاتب داخل البنك.

X_3 = نسبة القروض الخاصة مقارنةً بالقروض الكلية في البنك.

D_1 = 1 إذا كان تقييم الإدارة «جيد».

D_2 = 1 إذا كان تقييم الإدارة «مقبول جداً».

D_3 = 1 إذا كان تقييم الإدارة «مقبول»

D_4 = 1 إذا كان التحقيق تم مشاركةً مع حكومة الولاية.

الأرقام بين الأقواس هي الأخطاء القياسية المقدرة.

(**a**) فسر هذه النتائج.

(**b**) هل هناك أي مشكلة تواجهك في تيسير المتغيرات الوهمية في هذا النموذج خصوصاً وأن الـ Y موجودة في صورة لوغاريتم؟

(**c**) كيف يمكنك تفسير المتغيرات الوهمية؟

9.9 Sidney Langer، واحدة من طلابي، رغبت في معرفة أثر السياسة الفيدرالية في إعادة تقنين معدلات الفائدة بداية من يوليو 1979. لذلك قدرت النموذج التالي في الفترة الربع سنوية من III–1975 إلى II–1983(*):

$$\begin{aligned}\hat{Y}_t = \; & 8.5871 - 0.1328P_t - 0.7102\text{Un}_t - 0.2389M_t \\ \text{se}\; & (1.9563) \quad (0.0992) \quad (0.1909) \quad (0.0727) \\ & + 0.6592Y_{t-1} + 2.5831\text{Dum}_t \qquad R^2 = 0.9156 \\ & \quad (0.1036) \qquad (0.7549)\end{aligned}$$

حيث: Y = الفائدة على الشهادات الشهرية (كل 3 شهور).

P = معدل التضخم المتوقع.

Un = معدل البطالة المخلص من الأثر الموسمي.

M = تغيرات الأصول النقدية.

Dom = متغير وهمي يأخذ القيمة 1 للمشاهدات بداية من 1 يوليو 1979.

(**a**) فسر هذه النتائج.

(**b**) ما هو أثر إعادة تقنين معدل الفائدة؟ هل هذه النتائج لها أي مغزى اقتصادي؟

(*) "Sidney Langer, "Interest Rate Deregulation and Short-Term Interest Rates," unpublished term paper.

(c) معاملات P_t، Un_t و M_t سالبة، هل لديك تعليل اقتصادي لذلك؟

10.9 بالرجوع إلى الانحدار الجزيئي الذي ناقشناه من قبل. افترض بالإضافة إلى وجود تغير في معامل الميل عند X^* فهناك أيضاً قفزة في خط الانحدار، كما هو موضح في الشكل (7.9). كيف يمكنك تعديل (1.8.9) حتى تأخذ في الاعتبار القفزة الموجودة في خط الانحدار عند X^*؟

11.9 محدد سعر الكوكاكولا للأونس الواحد Cathyschaefer، إحدى طالباتي، قدرت الانحدار التالي من بيانات مقطعية من 77 مشاهدة: (*)

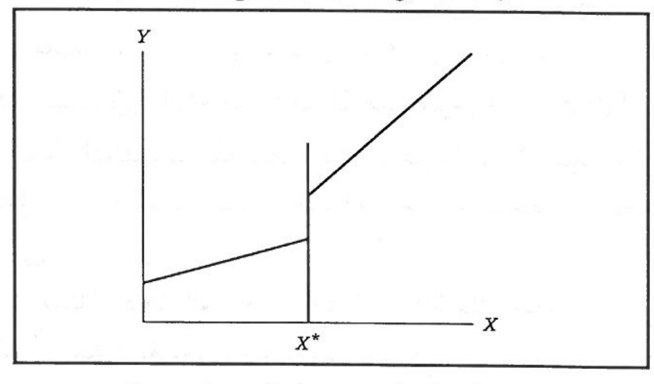
$$P_i = \beta_0 + \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \mu_i$$

حيث: P_i = سعر الكوكاكولا لكل أونس (oz)

D_{1i} = 001 إذا كان المتجر لبيع بقايا المنتجات والتي تباع بسعر مخفض.
= 010 إذا كان المتجر فرعاً لسلسلة متاجر كبيرة.
= 100 إذا كان المتجر صغيراً.

D_{2i} = 10 للسلع ذات العلامة التجارية.
= 01 للسلع بدون أي علامة تجارية.

D_{3i} = 0001 إذا كانت الزجاجة (2 لتر) أي 67.6 أونس (oz).
= 0010 إذا كانت الزجاجة من 28 إلى 33.8 أونس (oz).
(لاحظ أن 33.8 oz = 1 liter)
= 0100 إذا كانت الزجاجة 16 أونس (oz).
= 1000 إذا كان الزجاجة 16 أونس (oz).



شكل (7.9) انحدار خطي تجزيئي غير متصل

(*) Cathy Schaefer, " Price Per Ounce of Cola Beverage as a Function of Place of Purchase, Size of Container, and Branded or Unbranded Product," unpublished term project.

النتائج جاءت كالتالي:

$$\hat{P}_i = 0.0143 - 0.000004D_{1i} + 0.0090D_{2i} + 0.00001D_{3i}$$
$$Se = \quad (0.00001) \quad (0.00011) \quad (0.00000)$$
$$t = \quad (-0.3837) \quad (8.3927) \quad (5.8125)$$
$$R^2 = 0.6033$$

لاحظ أن: الأخطاء القياسية معطاة حتى العلامة العشرية الخامسة.

(a) علق على الأسلوب الذي تم به إدراج المتغيرات الوهمية في النموذج.

(b) بافتراض أن المتغيرات الوهمية وجدتها بهذا الشكل في النموذج سليمة كيف يمكنك تفسير النتائج.

(c) معامل D_3 موجب ومعنوي إحصائياً. كيف يمكنك تعليل هذه النتيجة؟

12.9 في دراسة متعلقة بالدخل بالنسبة للفرد مقاس بالدولار (X)، والعمر المتوقع بالسنوات (Y)، حصل Sen وSrivastara على نتائج الانحدار التالية مستخدمين بيانات من 101 دولة(*).

$$\hat{Y}_i = -2.40 + 9.39 \ln X_i - 3.36 [D_i(\ln X_i - 7)]$$
$$se = \quad (4.73) \quad (0.859) \quad (2.42) \qquad R^2 = 0.752$$

حيث: $D_i = 1$ إذا كان $\ln X_i > 7$

$= 0$ بخلاف ذلك

لاحظ أن $\ln X_i = 7$، $X = \$1097$ (تقريباً).

(a) ما هي الأسباب وراء استخدام متغير الدخل في الصورة اللوغاريتمية؟

(b) كيف يمكنك تفسير المعامل المساوي لـ 9.39 والخاص بـ $\ln X_i$؟

(c) ما هو السبب وراء استخدام المتغير المنحدر عليه $D_i(\ln X_i - 7)$؟

وكيف يمكنك تفسير هذا المتغير المنحدر عليه لغوياً؟ وكيف يمكنك تفسير المعامل 3.36- الخاص بهذا المتغير (ملاحظة مساعدة: الانحدار الخطي الجزيئي).

(d) إذا افترضنا أن دخل الفرد والمساوي لـ 1097$ يمثل الخط الفاصل بين البلاد الفقيرة والبلاد الغنية، كيف يمكنك استنتاج انحدار يمثل البلاد التي يكون

(*) Ashish Sen and Muni Srivastava, Regression Analysis: Theory, Methods, and Applications, Springer–Verlag, New York, 1990, p. 92.
مع ملاحظة أنه تم تغيير الرموز.

فيها دخل الفرد أقل من 1097$ ، وانحدار آخر للبلاد التي يكون فيها دخل الفرد أكثر من 1097$؟

(e) ما هي الاستنتاجات العامة التي يمكنك الوصول إليها من نتائج الانحدار الذي ناقشناه في هذا التمرين؟

13.9 اعتبر النموذج التالي :

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

حيث : $D_i = 0$ للعشرين مفردة الأولى .

$D_t = 1$ للثلاثين مفردة الباقية .

وإذا علمت أيضاً أن $\text{var}(u_i^2) = 300$

(a) كيف يمكنك تفسير الجزء الثابت β_1 وβ_2؟

(b) ماهي القيم المتوسط للمجموعتين؟

(c) كيف يمكنك حساب تباين $(\hat{\beta}_1 + \hat{\beta}_2)$؟ لاحظ أن : معطى لديك أن تغاير $(\hat{\beta}_1, \hat{\beta}_2)$ يساوي 15− أي أن $\text{cov}(\hat{\beta}_1, \hat{\beta}_2) = -15$

14.9 لدراسة أثر قوانين حقوق العمل على الاشتراك في الاتحادات العمالية بالولايات المتحدة (مع استثناء المناطق التي تشترط الاشتراك في هذه الاتحادات كخطوة أولية للعمل). تم التوصل لنتائج الانحدار التالية بناء على بيانات 50 ولاية من الولايات المتحدة الأمريكية في العام 1982 :(*)

$$\widehat{\text{PVT}}_i = 19.8066 - 9.3917\ \text{RTW}_i$$
$$t = (17.0352) \quad (-5.1086)$$
$$r^2 = 0.3522$$

حيث : PVT = نسبة العاملين في القطاع الخاص في الاتحادات العمالية عام 1982 و RTW = 1 إذا كانت قوانين حقوق العمل موجودة بالولاية 0 بخلاف ذلك .

لاحظ أنه في عام 1982 كانت هناك قوانين لحقوق العمل موجودة في 20 ولاية فقط .

(a) مبدئياً . . ما هي العلاقة المتوقعة بين PVT وRTW؟

(b) هل نتائج الانحدار مؤيدة للعلاقة المتوقعة السابقة؟

(*) البيانات المستخدمة في نتائج الانحدار تم الحصول عليها من :

N. M. Meltz, "Interstate and Interprovincial Differences in Union Density," Industrial Relatins, vol. 28, no. 2, 1989, pp. 142–158.

(c) فسر نتائج الانحدار.

(d) ما هو متوسط نسبة العاملين في القطاع الخاص، والموجودين في الاتحادات العمالية بالولايات التي ليس لديها قوانين لحقوق العمل؟

15.9 اعتبر نموذج الانحدار التالي:

$$Y_i = \beta_1 + \beta_2 D_i + u_i$$

حيث: Y تمثل الأجر بالساعة بالدولار، وD متغير وهمي يأخذ القيمة 1 لخريج الجامعة، و0 للحاصل على الشهادة الثانوية فقط. باستخدام معادلات الـ OLS المعطاة في الفصل (3)، اثبت أن $\hat{\beta}_1 = \bar{Y}_{hg}$ و $\hat{\beta}_2 = \bar{Y}_{cg} - \bar{Y}_{hg}$ حيث الرموز تعني التالي: hg = الحاصل على الشهادة الثانوية فقط وcg = الحاصل على شهادة جامعية، والقيمة تشتمل على n_1 مفردة من الحاصلين على الشهادة الثانوية فقط، و n_2 مفردة من الحاصلين على الشهادة الجامعية، وبالتالي إجمالي القيمة هو $n = n_1 + n_2$

16.9 لدراسة معدل النمو في البرازيل خلال الفترة 1970 إلى 1992، قام Mukherfec et al. بتقدير نموذج الانحدار التالي(*):

النموذج I:

$$\widehat{\ln(\text{Pop})}_t = 4.73 + 0.024t$$
$$t = (781.25) \quad (54.71)$$

النموذج II:

$$\widehat{\ln(\text{Pop})}_t = 4.77 + 0.015t - 0.075D_t + 0.011(D_t t)$$
$$t = (2477.92) \quad (34.01) \quad (-17.03) \quad (25.54)$$

حيث: PoP = عدد السكان بالمليون نسمة، t: متغير الاتجاه العام.

$D_t = 1$ للمشاهدات بداية من 1978، و0 قبل 1978، و ln ترمز إلى اللوغاريتم الطبيعي.

(a) في النموذج I، ما هو معدل النمو في البرازيل خلال فترة الدراسة؟

(b) هل هناك فرق إحصائي بين معدلات النمو قبل وبعد 1978؟ كيف عرفت ذلك؟ إذا كان هناك اختلاف، ما هو معدل النمو في الفترة 1972 – 1977 والفترة 1978 – 1992؟

(*) Chandan Mukherjee, Howard White, and Marc Wuyts, Econometrics and Data Analysis for Developing Countries, Routledge, London, 1998, pp. 372–375.

مع ملاحظة تعديل بعض الرموز.

مسـائل : Problems

17.9 باستخدام البيانات الموجودة في جدول (1.9)، اختبر الفرض القائل بأن تباين الأخطاء في الفترتين الجزيئيتين IV-19858 إلى III-1966 و IV-1966 إلى II-1971 متساويين.

18.9 باستخدام الأسلوب المشروع في الفصل (8)، قارن بين الانحدار المقيد وغير المقيد، (3.7.9) و(4.7.9)، أي اختبر مدى صحة الفروض المقترحة.

19.9 في انحدار الدخل - الادخار الخاص بالولايات المتحدة (4.5.9) والذي ناقشناه في هذا الفصل، افترض أنه بدلاً من استخدام 1 و 0 كمتغير وهمي، قمت باستخدام $Z_i = a + bD_i$، حيث 0 أو $D_i = 1$، $a = 2$ و$b = 3$. قارن النتائج.

20.9 باستخدام نفس الانحدار السابق (4.5.9) والخاص بالدخل والادخار افترض أنك استخدمت $D_i = 0$ للمفردات في الفترة الثانية $D_i = 1$ للمفردات الموجودة في الفترة الأولى. كيف ستتغير النتائج المعطاة في (4.5.9)؟

21.9 استخدم البيانات المعطاة في جدول (2.9) وأعتبر النموذج التالي:

$$\ln \text{Savings}_i = \beta_1 + \beta_3 \ \ln \text{Income}_i + \beta_3 \ln D_i + u_i$$

حيث : ln تعني أننا استخدمنا الدالة اللوغاريتمية الطبيعية، و$D_i = 1$ للفترة 1970 − 1981 و 10 للفترة 1982 − 1995.

(a) ما هو التبرير المتوقع لاستخدام المتغير الوهمي بالشكل السابق؟

(b) قدر النموذج السابق وفسر النتائج.

(c) ما هي قيم الأجزاء الثانية لدالة الادخار في الفترتين الجزيئيتين وكيف يمكن تفسيرهما؟

22.9 بالرجوع إلى بيانات مبيعات الأجهزة الكهربائية الربع سنوية المعطاة في جدول (3.9). اعتبر النموذج التالي:

$$\text{Sales}_i = \alpha_1 + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \alpha_4 D_{4i} + u_i$$

حيث إن الـ D's هي متغيرات وهمية تأخذ القيم 1، و 0 للفترات الربع سنوية II حتى IV.

(a) قدر النموذج السابق لكل من: غسالات الأطباق، ومفرمة البواقي، وغسالات الملابس، كل على حدة.

(b) كيف تفسر معاملات الميل المقدرة.

(c) كيف يمكنك استخدام القيم المقدرة للـ α's لتخليص بيانات المبيعات من الأثر الموسمي لكل من الأجهزة السابقة كلاً على حدة؟

23.9 أعد تقرير النموذج الموجودة في تمرين 22.9 بعد إضافة متغير الإنفاق على السلع المعمرة كمتغير منحدر عليه.

(a) هل هناك اختلاف في نتائج الانحدار عن نظيرها التي حصلت عليها من قبل في تمرين 22.9؟ وإذا كان ذلك متحققًا بالفعل ما هو تفسيرك لهذا الاختلاف؟

(b) هل هناك أثر موسمي في بيانات الإنفاق على السلع المعمرة؟ كيف يمكنك التعامل مع ذلك؟

24.9 جدول (8.9) يعطي بيانات عن ربع سنوية عن انتخابات الرئاسة في الولايات المتحدة خلال الفترة 1916 إلى 1996. (*)

(a) باستخدام البيانات الموجودة في جدول (6.9). كون نموذجًا مناسبًا للتنبؤ بنسبة التصويت للحزب الديمقراطي.

(b) كيف يمكنك استخدام هذا النموذج للتنبؤ بنتائج انتخابات الرئاسة؟

جدول (8.9) بيانات انتخابات الرئاسة الأمريكية، 1916 – 1996

Year	*V*	*W*	*D*	*G*	*I*	*N*	*P*
1916	0.5168	0	1	2.229	1	3	4.252
1920	0.3612	1	0	−11.463	1	5	16.535
1924	0.4176	0	−1	−3.872	−1	10	5.161
1928	0.4118	0	0	4.623	−1	7	0.183
1932	0.5916	0	−1	−14.901	−1	4	7.069
1936	0.6246	0	1	11.921	1	9	2.362
1940	0.5500	0	1	3.708	1	8	0.028
1944	0.5377	1	1	4.119	1	14	5.678
1948	0.5237	1	1	1.849	1	5	8.722
1952	0.4460	0	0	0.627	1	6	2.288
1956	0.4224	0	−1	−1.527	−1	5	1.936
1960	0.5009	0	0	0.114	−1	5	1.932
1964	0.6134	0	1	5.054	1	10	1.247
1968	0.4960	0	0	4.836	1	7	3.215
1972	0.3821	0	−1	6.278	−1	4	4.766
1976	0.5105	0	0	3.663	−1	4	7.657
1980	0.4470	0	1	−3.789	1	5	8.093
1984	0.4083	0	−1	5.387	−1	7	5.403
1988	0.4610	0	0	2.068	−1	6	3.272
1992	0.5345	0	−1	2.293	−1	1	3.692
1996	0.5474	0	1	2.918	1	3	2.268

(*) هذه البيانات تم الحصول عليها من Ray Fair في جامعة Yale، حيث قام بالتنبؤ بنتيجة انتخابات الرئاسة لسنوات عديدة سابقة. البيانات إعادة تقييمها في Samprit Chatterjee, Ali S. Hadi, and Petram Price, Regression Analysis by Example, 3d ed., John Wiley & Sons, New York, 2000, pp. 150–151.

ملاحظات :

year سنة الانتخاب .

V نصيب الحزب الديموقراطي من التصويت الرئاسي للحزبين .

I متغير وهمي (1 إذا كان هناك تفوق للحزب الديموقراطي في الانتخابات و -1 إذا كان هناك تفوق للحزب الجمهوري) .

D متغير وهمي (1 إذا كان الحزب الديموقراطي مشاركًا في الانتخابات ، -1 إذا كان الحزب الجمهوري هو المشارك ، 0 بخلاف ذلك) .

W متغير وهمي (1 لانتخابات عام 1920 ، 1944 و 1948 ، 0 بخلاف ذلك) .

G معدل النمو للفرد في الثلاثة أرباع الأولى للسنة الانتخابية .

P القيمة المطلقة لمعدل النمو الخاص بالـ GDP في الـ 10 ربع الأولى من السنة .

N عدد أرباح السنة في الـ 10 ربع الأولى من السنة التي يكون فيها معدل النمو الـ GDP بالنسبة للفرد أعلى من 3.2% .

(c) chatlerjee et al. اقترح النموذج التالي كنموذج تجريبي للتنبؤ بنتائج الانتخابات الرئاسية :

$$V = \beta_0 + \beta_1 I + \beta_2 D + \beta_3 W + \beta_4 (GI) + \beta_5 P + \beta_6 N + u$$

قدر النموذج وعلق على النتائج وعلاقتها بالنتائج التي حصلت عليها من النموذج المختار سابقاً .

25.9 بالعودة إلى انحدار (4.6.9) . اختبر الفرض القائل بأن معدل زيادة متوسط أجر الساعة مقارنةً بالمستوى التعليمي يختلف وفقاً للنوع والعرق . (ملاحظة : استخدم متغيرات وهمية ضربية) .

26.9 بالعودة إلى انحدار (1.3.9) . كيف يمكنك تعديل النموذج حتى تستطيع معرفة ما إذا كان هناك تفاعل بين النوع ومحل السكن؟ قدم نتائجك وفقاً لهذا النموذج وقارنها مع النتائج التي تم الحصول عليها في (1.3.9) .

27.9 في النموذج $Y_i = \beta_1 + \beta_2 D_i + u_i$ اجعل $D_i = 0$ للـ 40 مشاهدة الأولى و $D_i = 1$ للمشاهدات الـ 60 الباقية . إذا علمت أن u_i لها وسط حسابي يساوي الصفر ، وتباين يساوي 100 . ما هي قيم الوسط والتباين الخاصة بالمجموعتين السابقتين من المشاهدات؟(*)

28.9 بالعودة إلى انحدار الدخل - الادخار للولايات المتحدة ، والذي تم استعراضه من قبل في هذا الفصل . كبديل لـ (1.5.9) دعنا نعتبر النموذج التالي :

$$\ln Y_t = b_i + \beta_2 D_t + \beta_3 X_t + \beta_4 (D_t X_t) + u_t$$

حيث : Y = الادخار ، و X = الدخل .

(*) هذا المثال مأخوذ من : Peter Kennedy, A Guide to Econometrics, 4th ed., MIT Press, Cambridge, Mass., 1998, p. 347.

(a) قدر النموذج السابق ، وقارن الناتج مع المعطاة في (4.5.9). أي النموذجين أفضل؟

(b) كيف يمكنك تفسير معامل المتغير الوهمي في هذا النموذج؟

(c) كما سنرى لاحقاً في باب عدم ثبات التباين، عادة ما نستخدم تحويل اللوغاريتم للمتغير التابع في تقليل عدم ثبات التباين في البيانات. دعنا نتحقق من ذلك في المثال الحالي عن طريق إجراء تحويل اللوغاريتم للمتغير Y وانحداره على X في الفترتين السابقتين ونرى ما إذا كان تباين الأخطاء المقدرة في الفترتين متساويًا إحصائياً أم لا. إذا تساوى ممكن استخدام اختبار Chow لتجميع البيانات بالأسلوب السابق ذكره بالتفصيل في هذا الباب.

APPENDIX 9A ملحق A9

انحدار شبه لوغاريتمي ذو متغير منحدر وهمي

Semilogarithmic Regression with Dummy Regressor

في الفقرة (10.9) لاحظنا في النماذج التالية :

$$\ln Y_i = \beta_1 + \beta_2 D_i \qquad (1)$$

أن التغير النسبي في Y (أي المرونة النسبية) ، بالنسبة للمتغير المنحدر الوهمي الذي يأخذ القيم 0 أو 1 يمكن الحصول عليها (كمعكوس اللوغاريتم للقيمة المقدرة لـ β_2 وطرحها من 1) مضروب في 100 ، أي يمكن الحصول عليها كالتالي :

$$(e^{\hat{\beta}_2} - 1) \times 100 \qquad (2)$$

الاثبات كالتالي : بما أن ln و exp ($=e$) هما دوال عكسية ، يمكننا كتابة (1) كالتالي :

$$\ln Y_i = \beta_1 + \ln\left(e^{\beta_2 D_i}\right) \qquad (3)$$

والآن عندما يكون $D = 0$ ، $e^{\beta_2 D_i} = 1$ و $D = 1$ ، $e^{\beta_2 D_i} = e^{\beta_2}$. وبالتالي عند الانتقال من الحالة 0 إلى الحالة 1 فإن $\ln Y_i$ يتغير بمقدار يساوي $(e^{\beta_2} - 1)$. ولكن التغير في log المتغير هو تغير نسبي والذي يمكن ضربه في 100 حتى يصبح نسبة . وبالتالي التغير كنسبة يساوي $(e^{\beta_2} - 1) \times 100$ كما هو مطلوب (لاحظ أن : $\ln_e e = 1$ ، أي أن log e للأساس e يساوي 1 ، كما الحال في log 10 للأساس 10 والذي يساي 1 أيضًا . تذكر أن log للأساس e يسمى اللوغاريتم الطبيعي أما log للأساس 10 يسمى اللوغاريتم المعتاد) .

الفصل العاشر

تعدد العلاقات الخطية: ماذا يحدث إذا كانت المتغيرات المنحدرة مرتبطة؟

MULTICOLLINEARITY: WHAT HAPPENS IF THE REGRESSORS ARE CORRELATED?

لا يوجد زوج من الكلمات تم استخدامه بشكل سيء، سواء في الاقتصاد القياسي أو التطبيقي، أكثر من (مشكلة تعدد العلاقات الخطية). حيث إن وجود عدد من المتغيرات المفسرة والمرتبطة خطيًا بدرجة كبيرة يعتبر أمرًا متكرر الحدوث في الواقع العملي.

ومن المعروف أن تصميم التجربة بناء على (مصفوفة البيانات)X'X التي تحتوي على بيانات حقيقية من الواقع (أي العينة محل الدراسة) يكون أفضل كثيراً للدراسة العملية، ولكن تظهر المشكلة السابقة من واقعية البيانات، ويصبح تصميم التجربة سيئًا، ويصبح الانحدار التدريجي وانحدار الحافة غير مناسبين تمامًا للبيانات. ويجب علينا أن نتقبل الحقيقة الخاصة بأن البيانات التي نحصل عليها بدون تصميم للتجربة (أي البيانات المجمعة بدون سابق تصميم للتجربة) ليست دائمًا ممثلة بشكل جيد للمعالم محل الدراسة[1].

الافتراض العاشر في نماذج الانحدار الخطي التقليدي (CLRM) ينص على عدم وجود تعدد للعلاقات الخطية بين المتغيرات المنحدرة المتضمنة في نموذج الانحدار. **في هذا الفصل، سنستعرض هذا الفرض بمزيد من التفصيل، وسنحاول الإجابة على الأسئلة التالية:**

(1) Edward E. Leamer, "model Choice and Specification Analysis," in Zvi Griliches and Michael D. Interiligator, eds., Handbook of Econometrics, vol. I, North Holland Publishing Company, Amsterdam, 1983, pp. 300-301.

1 - ما هي طبيعة مشكلة تعدد العلاقات الخطية؟

2 - هل تعدد العلاقات الخطية يعتبر بالفعل مشكلة؟

3 - ما هي العواقب العملية لوجود ظاهرة تعدد العلاقات الخطية؟

4 - كيف يمكن اكتشافها؟

5 - ما هي الأساليب التي يمكن استخدامها لعلاج أو التخفيف من مشكلة تعدد العلاقات الخطية؟

في هذا الفصل أيضًا، سنناقش الفرض 7 للـ CLRM والخاص بأن عدد المشاهدات في العينة يجب أن يكون أكبر من عدد المتغيرات المنحدرة. وأيضًا سنناقش الفرض 8، والذي يتطلب وجود قدر ما من التباين في قيم المتغيرات المنحدرة، وهذا الفرض متعلق بشكل مباشر بفرض تعدد العلاقات الخطية Arthur Goldberger أطلق على الفرض 7 مشكلة التصغير الجزيئي [2] والمقصود به ببساطة هو الحجم الصغير للعينة.

1.10 طبيعة تعدد العلاقات الخطية
THE NATURE OF MULTICOLLINEARITY

مصطلح تعدد العلاقات الخطية يعود إلى Ragnar Frisch [3].

أصلاً هذا المصطلح يعني وجود علاقة خطية تامة وكاملة بين بعض أو كل المتغيرات المفسرة الموجودة في نموذج الانحدار[4]. في نموذج الانحدار الذي يحتوي على k من المتغيرات المفسرة X_1، X_2، ... ، X_k (حيث $X_1 = 1$ لكل المشاهدات، حيث نعبر عن الجزء الثابت المقطوع من المحور الصادي)، العلاقة الخطية التامة تكون موجودة ومتحققة في هذا النموذج إذا توفر الشرط التالي:

$$\lambda_1 X_1 + \lambda_2 X_2 + \dots + \lambda_k X_k = 0 \qquad (1.1.10)$$

(2) انظر في محاضرات الاقتصاد القياسي لـ Harvard University Press, Cambridge, Mass., 1991, p. 249.

(3) Ragnar Frisch, Statistcal Confluence Analysis by Means of Complete Regression Systems, Institute of Economics, Oslo University, publ, no. 5, 1934.

(4) بشكل أكثر تحديداً، تعدد العلاقات الخطية يعني أكثر من واحدة من العلاقات الخطية، أما العلاقات الخطية فتسعى وجود علاقة واحدة خطية فقط. ولكن هذه التفرقة غير موجودة في الواقع العملي، ويتم استخدام تعدد العلاقات الخطية للإشارة إلى الحالتين السابقتين معاً.

حيث λ_1، λ_2، ... ، λ_k ثوابت، بحيث لا يساوي الصفر جميعًا بشكل آني[5].

وبوجه عام، يتم الآن استخدام مصطلح تعدد العلاقات الخطية للتعبير عن حالة التعدد الكامل في العلاقات الخطية، كما هو موضح في (1.1.10) بالإضافة إلى الحالة التي تكون فيها المتغيرات المقدرة Xi's مرتبطة، ولكن ليست بشكل كامل، كما هو موضح كالتالي[6]:

$$\lambda_1 X_1 + \lambda_2 X_2 + \ldots + \lambda_2 X_k + v_i = 0 \qquad (2.1.10)$$

حيث v_i يمثل حد الخطأ العشوائي.

لمعرفة الفرق بين تعدد العلاقات الخطية الكاملة والأقل من الكاملة، دعنا على سبيل المثال نفترض أن $\lambda_2 \neq 0$، وبالتالي فإن (1.1.10) يمكن إعادة كتابتها كالتالي:

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \cdots - \frac{\lambda_k}{\lambda_2} X_{ki} \qquad (3.1.10)$$

المعادلة السابقة توضح العلاقة الخطية التامة بين X_2 وباقي المتغيرات، وتوضح أيضًا كيفية اشتقاق X_2 كعلاقة خطية في المتغيرات المفسرة الأخرى. في هذه الحالة، معامل الارتباط بين المتغير X_2 التوليفة الخطية الموجودة في الجانب الأيمن من (3.1.10) محدود بالواحد الصحيح.

وبالمثل إذا كان $\lambda_2 \neq 0$ فإن المعادلة (2.1.10) يمكن كتابتها كالتالي:

$$X_{2i} = -\frac{\lambda_1}{\lambda_2} X_{1i} - \frac{\lambda_3}{\lambda_2} X_{3i} - \cdots - \frac{\lambda_k}{\lambda_2} X_{ki} - \frac{1}{\lambda_2} v_i \qquad (4.1.10)$$

وهي توضح أن X_2 ليس توليفة خطية تامة من باقي X's حيث إنه يوجد مقدار من الخطأ العشوائي v_i.

كمثال رقمي، دعنا نستعرض البيانات الافتراضية التالية:

(5) احتمال حصول الفرد على عينة بها قيم التنبؤات المنحدرة مرتبطة بهذا الشكل صغير جدًا في الناحية العملية إلا إذا كان هناك تصميم ما للتجربة مثلاً يكون عدد المشاهدات أقل من عدد المتغيرات المنحدرة أو يقع الفرد في مصيدة المتغيرات الوهمية التي ناقشناها من قبل في (الفصل 9) انظر تمرين 2.10.

(6) إذا كان هناك متغيران مفسران مرتبطان، فإنه يمكن قياس هذا الارتباط باستخدام معامل الارتباط البسيط، ولكن إذا كان هناك أكثر من متغيرين اثنين فإنه يمكن قياس الارتباط بمعاملات الارتباط الجزيئية أو معامل الارتباط المتعدد R لمتغير واحد X مع باقي المتغيرات المنحدرة معاً.

X_2	X_3	X_3^*
10	50	52
15	75	75
18	90	97
24	120	129
30	150	152

من الواضح أن $X_{3i} = 5X_{2i}$. وبالتالي فإنه توجد علاقة متعددة تامة بين X_2 و X_3، حيث إن معامل الارتباط r_{23} يساوي الواحد الصحيح.

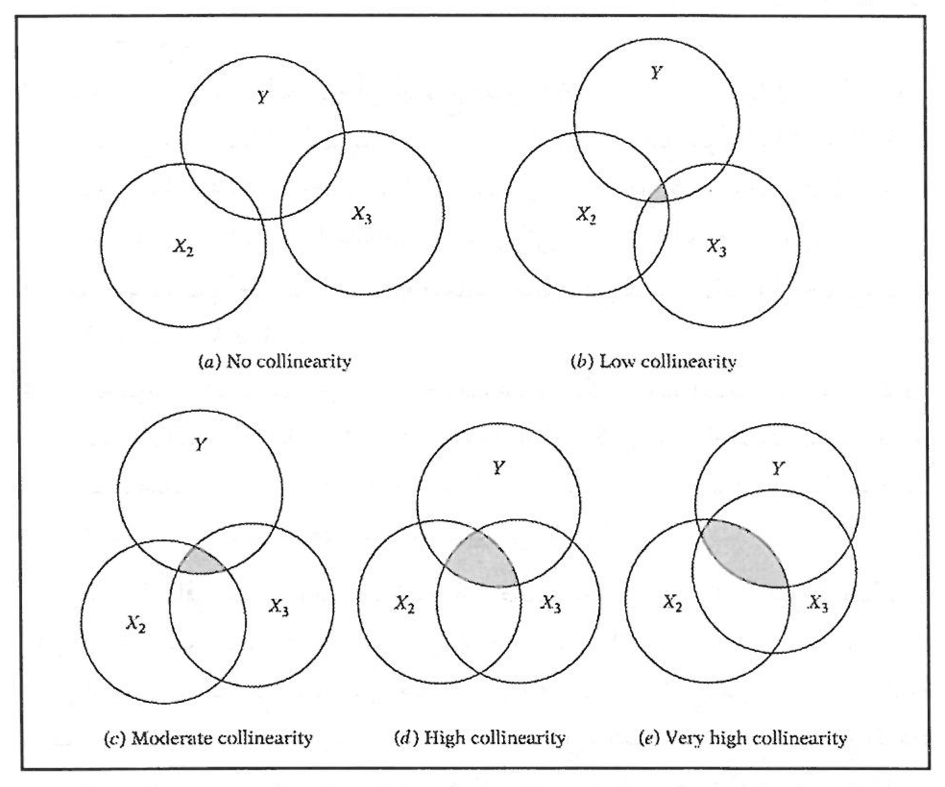
المتغير X_3^* تم تكوينه من X_3 بإضافة الأرقام التالية – والتي تم الحصول عليها من جدول الأرقام العشوائية – 2، 0، 7، 9، 2. والآن لم يعد هناك ارتباط متعدد كامل بين X_2 و X_3^*. عمومًا المتغيران مرتبطان بدرجة عالية، حيث إنه من حساب معامل الارتباط بينهما وجد أنه يساوي 0.9959.

الأسلوب الجبرى السابق للتعبير عن تعدد العلاقات الخطية يمكن التعبير عنه بيانيًا باستخدام Ballentine (ارجع إلى الشكل (3.9)، كرر الحصول على الشكل (1.10)). في هذا الشكل البياني الدوائر Y، X_2 و X_3 تمثل على التوالي التباين في Y (المتغير التابع) و X_2 وX_3 (المتغيرات المفسرة). درجة تعدد العلاقات الخطية يمكن قياسها بمقدار التداخل (المنطقة المظللة) من دوائر X_2 وX_3. في شكل (a1.10) لا يوجد أي تداخل بين X_2 وX_3 وبالتالي لا يوجد تعدد في العلاقات الخطية بينهما. في الأشكال من (b1.10) إلى (e1.10) هناك درجات من تعدد العلاقات الخطية تتعاون من درجات ضعيفة إلى قوية – كلما زادت المنطقة المتداخلة بين X_2، X_3 (أي كلما زادت المنطقة المظللة) كلما كان ذلك دليلاً على زيادة درجة تعدد العلاقة الخطية بين المتغيرين. الحالة القصوى عندما يكون هناك تداخل كامل بين X_2 و X_3 (أو تكون X_2 بالكامل داخل X_3 أو العكس) تسمى تعدد علاقات خطية كامل.

كملحوظة عابرة، لاحظ أن تعدد العلاقات الخطية – كما سبق وعرفناه – يشير إلى علاقات خطية فقط بين المتغيرات المفسرة. ولا يعبر على الإطلاق عن العلاقات غير الخطية بينهم. فعلى سبيل المثال، اعتبر نموذج الانحدار التالي:

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i \qquad \textbf{(5.1.10)}$$

حيث Y = التكلفة الكلية للإنتاج، X = الناتج. المتغيرات X_i^2 (تربيع الناجح) و X_i^3 (تكعيب الناجح) مرتبطين بشكل دالي بالمتغير X. ولكن العلاقة غير خطية. وبالتالي فإن النماذج المماثلة للنموذج (5.1.10) لا تخالف فرض عدم وجود تعدد في العلاقات الخطية. عمومًا كمشكلة واقعية، فإن معاملات الارتباط المقاسة ستظهر أن X_i، X_i^2 و X_i^3 مرتبطة بدرجة عالية، والذي سيجعل هناك صعوبة في تقدير المعاملات الخاصة بـ (5.1.10) كما سنوضح لاحقًا بدرجة تقدير عالية (أي بأخطاء قياسية صغيرة).



شكل (1.10) معاملات الارتباط المقاسة مرتبطة بدرجة عالية

لماذا يفترض نموذج الانحدار الخطي التقليدي عدم وجود تعدد في العلاقات الخطية بين المتغيرات المفسرة X's؟ السبب في أن: إذا كان التعدد في العلاقات الخطية كامل كما في (1.1.10) فإن معاملات الانحدار للمتغيرات X غير محددة، وأخطاءهم القياسية لا نهائية (غير محدودة)، إذا كان تعدد العلاقات الخطية ليس كاملاً تمامًا كما

في (2.1.10) فإن معاملات الانحدار على الرغم من امكانية تقديرها، إلا أن هذا التقدير يكون مصاحبًا لأخطاء قياسية عالية (مقارنة بقيم المعاملات نفسها) مما يعني أن المعاملات لا يمكن تقديرها بدقة عالية.

هذه العبارات موضحة تفصيليلاً في الفقرات التالية. هناك مصادر عديدة لتعدد العلاقات الخطية. كما ذكر Peck و Montgomery **فإن تعدد العلاقات الخطية يعود إلى العوامل التالية:**[7]

1 - أسلوب جمع البيانات: فمثلاً أخذ عينة على قيم محدودة للمتغيرات المنحدرة في المجتمع.

2 - وجود قيود على النموذج أو المجتمع الذي تؤخذ منه العينة: فمثلاً في الانحدار الخاص باستهلاك الكهرباء على الدخل (X_2) وحجم المنزل (X_3). هناك قيد عملي على المجتمع، حيث إن العائلات التي لها دخل كبير عادة تكون أحجام منازلها كبيرة وأكبر من العائلات الأخرى التي من ذوات الدخل المحدود.

3 - توصيف النموذج: فمثلاً إضافة متعددة حدود إلى نموذج الانحدار خصوصًا إذا كان مدى الـ X صغير.

4 - التحديد الزائد للنموذج: هذا يحدث عندما يكون عدد المتغيرات المفسرة في النموذج أكبر من عدد المشاهدات، هذا يحدث كثيرًا في المجالات الطبية عندما يكون هناك عدد محدود أو صغير من المرضى يصاحبهم عدد كبير من المعلومات والبيانات التي يتم جمعها عن عدد كبير من المتغيرات.

سبب إضافي **لتعدد العلاقات الخطية**، خصوصًا في مجال السلاسل الزمنية، أن تكون المتغيرات المنحدرة متقاسمة في اتجاه عام مشترك، أي أنهم يزدادون أو يقلون بمرور الزمن. فمثلاً في الانحدار الخاص بالإنفاق الاستهلاكي على الدخل، الثروة وحجم المجتمع، فإن المتغيرات المنحدرة (الدخل، الثروة، حجم المجتمع) تزداد جميعها بمرور الزمن بمعدل متساو أو مختلف مما يؤدي إلى وجود تعدد في العلاقات الخطية بين هذه المتغيرات.

---

(7) Douglas Montgomery and Elizabeth Peck, Introduction to Linear Regression Analysis, John Wiley & Sons, New York, 1982, pp. 289-290. See also R. L. Mason, R. F. Gunst, and J. T. Webster, "Regression Analysis and Problems of Multicollinearity," Communications in Statistics A, vol. 4, no. 3 1975 pp. 277-292 R. F. Gunst and R. L. Mason "Advantages of Examining Multicollinearities in Regression Analysis," Biometrics, vol. 33, 1977, pp. 249 260.

2.10 التقدير في ظل وجود تعدد كامل للعلاقات الخطية: ESTIMATION IN THE PRESENCE OF PERFECT MULTICOLLINEARITY

سبق وأن ذكرنا أنه في حالة وجود تعدد كامل للعلاقات الخطية بين معاملات الانحدار، فإنها تصبح غير قابلة للتقدير وأخطاءهم القياسية غير محدودة. يمكن التحقق بشكل عملي من هذه الحقيقة باستخدام نموذج يحتوي على ثلاثة متغيرات مفسرة، وباستخدام الشكل الانحرافي (أي أن المتغيرات كلها يتم التعبير عنها في صورة انحرافتها عن الوسط الحسابي) يمكن كتابة نموذج الانحدار كالتالي:

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{u}_i \quad (1.2.10)$$

في الفصل 7 نجد أن

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2} \quad (7.4.7)$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i}x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right) - \left(\sum x_{2i}x_{3i}\right)^2} \quad (7.4.8)$$

افترض أن $X_{3i} = \lambda X_{2i}$، حيث λ ثابت غير صفري (مثلاً 2، 4، 1.8، ...) بالتعويض عن ذلك في (7.4.7) نحصل على

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\lambda^2 \sum x_{2i}^2\right) - \left(\lambda \sum y_i x_{2i}\right)\left(\lambda \sum x_{2i}^2\right)}{\left(\sum x_{2i}^2\right)\left(\lambda^2 \sum x_{2i}^2\right) - \lambda^2 \left(\sum x_{2i}^2\right)^2} = \frac{0}{0} \quad (2.2.10)$$

والذي يعتبر مقداراً غير معرف. يمكن للقارئ أن يثبت أيضًا أن $\hat{\beta}_3$ غير محدد.[8]

لماذا حصلنا على النتيجة الموجودة في (2.2.10)؟ تذكر معنى $\hat{\beta}_2$: إنه يمثل معدل التغير في متوسط قيمة Y عندما X_2 تتغير بمقدار الوحدة، مع افتراض ثبات X_3. ولكن إذا كانت X_2 و X_3 مرتبطين خطيًا بشكل تام، فليس هناك طريقة لافتراض ثبات X_3 مع

(8) طريقة أخرى لإثبات ذلك كالتالي: من التعريف فإن معامل الارتباطين X_2 و X_3 و r_{23}، يساوي $\sum x_{2i}x_{3i} / \sqrt{\sum x_{2i}^2 \sum x_{3i}^2}$. إذا كان r_{23}^2، أي هناك ارتباط متعدد تام بين X_2 و X_3 فإن مقام (7.4.7) سيساوي الصفر، وبالتالي يصبح تقدير β_2 (أو β_3) غير ممكن.

تغيير X_2 لأن X_3 تتغير أيضًا بالقيمة λ. وبالتالي، فإن المقصود بهذا هو عدم إمكانية التفرق بين تأثير X_2 و X_3 من العينة المختارة، فالأسباب عملية X_2 و X_3 لا يمكن التفرقة بينهما. في الاقتصاد القياسي التطبيقي هذه المشكلة تفسد الدراسة، حيث يكون الهدف الرئيسي هو فصل التأثير الجزيئي لكل متغير من المتغيرات المفسرة على المتغير التابع.

للتعبير عن ذلك بشكل آخر، دعنا نعوض عن $X_{3i} = \lambda X_{2i}$ في (1.2.10) وسنحصل على التالي [انظر أيضًا (9.1.7)] :

$$\begin{aligned} y_i &= \hat{\beta}_2 x_{2i} + \hat{\beta}_3(\lambda x_{2i}) + \hat{u}_i \\ &= (\hat{\beta}_2 + \lambda\hat{\beta}_3)x_{2i} + \hat{u}_i \\ &= \hat{\alpha} x_{2i} + \hat{u}_i \end{aligned} \tag{3.2.10}$$

حيث

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) \tag{4.2.10}$$

باستخدام معادلات OLS التقليدية على (3.2.10) نحصل على

$$\hat{\alpha} = (\hat{\beta}_2 + \lambda\hat{\beta}_3) = \frac{\sum x_{2i} y_i}{\sum x_{2i}^2} \tag{5.2.10}$$

وبالتالي، فإنه على الرغم من إمكانية تقدير α بشكل منفرد، فإنه لا توجد طريقة لتقدير β_2 و β_3 بشكل منفرد، رياضيًا :

$$\hat{\alpha} = \hat{\beta}_2 + \lambda\hat{\beta}_3 \tag{6.2.10}$$

يعطينا معادلة واحدة في مجهولين اثنين (لاحظ أن λ معطاة)، وبالتالي يوجد عدد لا نهائي من الحلول لـ (6.2.10) بناء على قيم معطاة لـ $\hat{\alpha}$ و λ.

للتعبير عن الفكرة بشكل محدد، دع $\hat{\alpha}$= 0.8 و 2 = λ وبالتالي يكون لدينا :

$$0.8 = \hat{\beta}_2 + 2\hat{\beta}_3 \tag{7.2.10}$$

أو

$$\hat{\beta}_2 = 0.8 - 2\hat{\beta}_3 \tag{8.2.10}$$

والآن اختر أي قيمة لـ $\hat{\beta}_3$، وسيكون لديك حل مباشر لـ $\hat{\beta}_2$. وإذا اخترت قيمة أخرى لـ $\hat{\beta}_3$ سيكون لديك حل آخر لـ $\hat{\beta}_2$. ومهما كررت المحاولة لن يكون لديك حل وحيد لـ $\hat{\beta}_2$.

خلاصة المناقشة السابقة، أنه في حالة تعدد العلاقات الخطية التام، يمكن أن نحصل على حل وحيد لمعاملات الانحدار المنفردة. ولكن لاحظ أننا نحصل على هذا الحل للتوليفة الخطية الموجودة في هذه المتغيرات.

التوليفة الخطية $(\beta_2 + \lambda\beta_3)$ يمكن تقديرها بشكل وحيد عن طريق α وقيمة معطاة لـ λ.[9]

وكملاحظة عابرة، نجد أن حالة تعدد العلاقات الخطية الكاملة أو التامة تكون المتغيرات والأخطاء القياسية لـ $\hat{\beta}_2$ و$\hat{\beta}_3$ غير محدودة. (انظر تمرين 21.10).

3.10 التقدير في ظل وجود تعدد في العلاقات الخطية "كبير" ولكن "غير كامل": ESTIMATION IN THE PRESENCE OF "HIGH" BUT "IMPERFECT" MULTICOLLINEARITY

التعدد الكامل في العلاقات الخطية يعتبر حالة شاذة. فبوجه عام لا توجد علاقة خطية تامة بين المتغيرات المفسرة خصوصًا عندما تتحلق البيانات بسلسلة زمنية اقتصادية. لذلك فمن الأفضل مناقشة النموذج ثلاثي المتغيرات في الشكل الانحرافي المعطي في (1.2.10) بدلاً من تعدد العلاقات الخطية الكامل. افترض أن لدينا التالي:

$$x_{3i} = \lambda x_{2i} + v_i \qquad (1.3.10)$$

حيث $\lambda \neq 0$ و v_i خطأ عشوائي، بحيث إن $\sum x_{2i}v_i = 0$ (لماذا؟)

يمكن أن نرى من خلال الشكل (b1.10) إلى الشكل (e1.10) ان لدينا حالة تعدد في العلاقات الخطية غير كاملة.

في هذه الحالة، تقدير معاملات الانحدار β_2 وβ_3 يكون ممكنًا. فعلى سبيل المثال، التعويض بـ (1.3.10) في (7.4.7) نحصل على

$$\hat{\beta}_2 = \frac{\sum(y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum y_i x_{2i} + \sum y_i v_i)(\lambda \sum x_{2i}^2)}{\sum x_{2i}^2 (\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum x_{2i}^2)^2} \qquad (2.3.10)$$

تحت شرط استخدام $\sum x_{2i}v_i = 0$. وبنفس الطريقة يمكن الحصول على $\hat{\beta}_3$.

(9) في أدبيات الاقتصاد القياسي، دالة مثل $(\beta_2 + \lambda\beta_3)$ تعرف على أنها دالة قابلة للتقدير.

الآن وعلى خلاف (2.2.10) لا يوجد سبب يجعلنا نعتقد مسبقًا أن (2.3.10) غير ممكن تقديرها. بالطبع إذا كان v_i صغيراً بشكل كاف، مثلاً، قريب جداً من العنصر فإن (1.3.10) ستكون تقريبًا مؤشرًا عن حالة تعدد كامل في العلاقات الخطية. وسنعود مرة أخرى إلى حالة عدم القدرة على التقدير الموجودة في (2.2.10).

4.10 تعدد العلاقات الخطية: ماهي صعوبات عدم المعرفة؟ العواقب النظرية لتعدد العلاقات الخطية

MULTICOLLINEARITY: MUCH ADO ABOUT NOTHING? THEORETICAL CONSEQUENCES OF MULTICOLLINEARITY

تذكر أنه إذا كانت فروض النموذج التقليدي متحققة، فإن مقدرات الـ OLS لمعاملات الانحدار تكون BLUE (أو BUE إذا أضفنا شرط الاعتيادية) الآن من الممكن إثبات أنه حتى إذا كان هناك تعدد في العلاقات الخطية يقترب من الحالة السابق ذكرها، فإن مقدرات الـ OLS مازالت تحتفظ بخاصية الـ BLUE [10]. لماذا إذن هناك قلق بخصوص تعدد العلاقات الخطية؟ وكما استعرض Christopher Achen (وهو أيضًا صاحب عبارات الافتتاحية الخاصة بهذا الفصل).

بعض الطلاب المبتدئين يقلقون بعض الشيء من وجود ارتباط بين المتغيرات المستقلة - وهذه المشكلة معروفة باسم مشكلة تعدد العلاقات الخطية. ولكن تعدد العلاقات الخطية لا يخالف أيًا من فروض الانحدار. فالمقدرات مازالت غير متحيزة ومتسقة، وأخطاؤها القياسية مقدرة بشكل سليم. الأثر الوحيد المرتب على تعدد العلاقات الخطية هو صعوبة الحصول على معاملات مقدرة بأخطاء قياسية صغيرة. ولكن إذا كان لدينا عدد قليل من المشاهدات، فإن هذا الأثر يحدث أيضًا فيكون لدينا متغيرات مستقلة بتباينات صغيرة (في الحقيقة، على المستوى النظري، فإن تعدد العلاقات الخطية وعدد المشاهدات القليل والتباين الصغير للمتغيرات المستقلة تعتبر جميعًا مشكلة واحدة) وبالتالي "ما الذي يمكن فعله عند وجود مشكلة تعدد العلاقات الخطية؟" هو سؤال مماثل للسؤال التالي "ما الذي يمكن فعله إذا لم يكن لدىً عدد كاف من المشاهدات؟". لا توجد إجابة إحصائية وافية لهذين السؤالين. [11]

---

(10) حيث إن الاقتراب من حالة تعدد العلاقات الخطية لا يتعارض مع الفروض السابق ذكرها في الفصل 7 فإن مقدرات OLS لا تزال BLUE كما هي.

(11) Christopher H. Achen, Interpreting and Using Regression, Sage Publications, Beverly Hills, Calif., 1982, pp. 82-83.

لإظهار أهمية حجم العينة، استخدم Goldberger مصطلح التصغير الجزيئي ليكون مناظر لمصطلح تعدد العلاقات الخطية. فوفقًا لـ Goldberger فإن التصغير الجزيئي الكامل (والذي يعتبر مناظر للتعدد الكامل للعلاقات الخطية) يحدث عندما تكون n، حجم العينة، تساوي الصفر. وفي مثل هذه الحالة، يكون التقدير غير ممكن. الحالة القريبة من التصغير الجزيئي، وهي المناظرة للحالة القريبة من تعدد العلاقات الخطية يكون عدد المشاهدات يزيد بالكاد عن عدد المعالم المراد تقديرها.

يرجع الفضل إلى Goldberger و Achen و Leamer لإلقاء الضوء على عدم اهتمام الآخرين بمشكلة حجم العينة وعدم الاهتمام بمشكلة تعدد العلاقات الخطية. للأسف، في العديد من الأبحاث التطبيقية التي تحتوى على بيانات ثانوية (أي بيانات تم تجميعها بواسطة بعض الأجهزة، مثل بيانات GNP المجمعة عن طريق الحكومة)، فإن الباحث قد لا يستطيع أن يفعل الكثير بخصوص حجم العينة، وقد يواجه " مشاكل في التقدير على قدر من الأهمية يجعلنا نتعامل معها [مشكلة تعدد العلاقات الخطية] كمخالف لنموذج CLR [الانحدار الخطي التقليدي classical linear regression] (12).

أولاً: دعنا نوضح التالي، أنه في حالة الاقتراب من تعدد العلاقات الخطية، فإن مقدرات OLS مازالت مقدرات غير متحيزة. ولكن عدم التحيز هو خاصية مرتبطة بالعينات المتكررة. المقصود أنه إذا افترضنا ثبات المتغيرات X وإذا حصلنا على عينات متكررة وحسبنا مقدرات الـ OLS لهذه العينات، فالمفترض أن يكون متوسط هذه القيم يؤول إلى قيم المجتمع الحقيقية لهذه المقدرات مع زيادة أحجام العينات. ولكن ذلك لا يعطينا أي معلومة عن خصائص التقدير في عينة ما محددة.

ثانيًا: أنه من المعلوم أن العلاقات الخطية لا تؤثر على خاصية التباين الأقل، حيث إنه في إطار المقدرات الخطية غير المتحيزة، فإن مقدرات OLS لها التباين الأقل أي أنها كفء. ولكن هذا لا يعني أن تباين مقدر الـ OLS سيكون بالضرورة صغيراً (في علاقته مع قيمة المقدر نفسه) عند تحديد عينة ما معطاة، وسنستعرض ذلك بالتفصيل لاحقًا.

(12) Peter Kennedy, A Guide to Econometrics, 3d ed., The MIT Press, Cambridge, Mass., 1992 p.177.

ثالثًا: تعدد العلاقات الخطية هو بالأساس خاصية مرتبطة بانحدار العينة، بمعنى أنه حتى إذا كانت المتغيرات المستقلة X ليست مرتبطة خطيًا في المجتمع قد يكونون مرتبطين بشكل واضح في بيانات عينة ما معطاة. فنحن عند التعامل مع دالة انحدار المجتمع أو الدالة النظرية (PRF) فإننا نعتقد أن المتغيرات X المتضمنة في النموذج لها تأثير منفصل ومستقل على المتغير التابع Y. لكن قد يحدث أنه في عينة ما معطاة والتي نستخدمها لاختيار PRF نجد أن بعض أو كل المتغيرات X مرتبطة بعلاقات خطية متعددة.

وبالتالي لا تستطيع فصل تأثيرها المنفرد على Y.

ولتوضيح الأمر، فإن العينة هي التي تخذلنا، فعلى الرغم من أن النظرية تجعلنا نعتقد أن كل الـ X's مهمة لتفسير المتغير Y فإن العينة قد لا تكون " غنية " بالشكل الكافي الذي يجعلنا نتضمن جميع المتغيرات المفسرة اللازمة في التحليل.

للتوضيح، دعنا نعود إلى مثال الدخل - الاستهلاك المستخدم من قبل في الفصل 3. الاقتصاديون يحددون نظريًا إلى جانب الدخل متغير آخر وهو ثروة المستهلك كمحدد مهم للإنفاق الاستهلاكي، وبالتالي فإنه يمكننا كتابة التالي:

$$\text{Consumption}_i = \beta_1 + \beta_2 \text{ Income}_i + \beta_3 \text{ Wealth}_i + u_i$$

الآن قد يحدث أنه عند حصولنا على بيانات عن الدخل والثروة، نجد أن المتغيرين مرتبطان بشكل كبير إن لم يكن كاملاً أو تامًا: فالأشخاص الأكثر ثراءً يفترض أن يكون لهم دخول أكبر. وبالتالي فعلى الرغم من أنه في النظرية الاقتصادية الدخل والثروة عاملان مفسران للسلوك الاستهلاكي، إلا أنه عمليًا (أي من خلال العينة) قد يكون من الصعب التفرقة بين تأثير كل من الدخل والثروة على النفقات الاستهلاكية للفرد.

وكحل نموذجي، فإنه لدراسة أثر كل من الدخل والثروة على الاستهلاك، فإننا نحتاج إلى حجم كاف من المشاهدات الثرية ذات الدخل المنخفض، ومشاهدات أخرى ذات دخل مرتفع وثروة محدودة (ارجع إلى الفرض 8). قد يكون ذلك متاحًا من خلال بيانات مقطعية (عن طريق زيادة حجم العينة) ولكنه شديد الصعوبة في الحدوث عندما يتعلق الأمر ببيانات سلاسل زمنية مجمعة.

لكل هذه الأسباب فحقيقة أن مقدرات الـ OLS وكونها مازالت BLUE رغم وجود مشكلة تعدد العلاقات الخطية ليس بالعزاء الكافي في الواقع العملي. فلابد من معرفة ما يحدث أو ما قد يحدث في إطار عينة ما محددة ومعطاة، وسيتم استعراض ذلك تفصيليًا في الفقرة التالية.

5.10 العواقب العملية لتعدد العلاقات الخطية: PRACTICAL CONSEQUENCES OF MULTICOLLINEARITY

في حالة تعدد العلاقات الخطية ، يجب أن يتوقع الفرد العواقب التالية :

1 - على الرغم من أن مقدرات OLS مازالت BLUE ، إلا أن لها تباينًا وتغايرًا كبيرًا مما يعيق جودة التقدير.

2 - يترتب على العاقبة الأولى، أن تكون فترات الثقة أوسع، مما يجعلنا نقبل الفرض العدمي الصفري (أي أن معامل المجتمع يساوي الصفر) بشكل كبير.

3 - أيضًا كنتيجة للعاقبة الأولى، فإن نتيجة اختيار t لواحد أو أكثر من المعاملات ستكون عادة غير معنوية.

4 - على الرغم من أن النسبة t لواحد أو أكثر من المعاملات ستكون غير معنوية إحصائيًا فإن، R^2، التي تمثل المقياس الخاص بجودة التوفيق، ستكون مرتفعة جدًا.

5 - مقدرات OLS وأخطاؤها القياسية ستكون حساسة جدًا لأي تغيير بسيط في البيانات.

العواقب السابقة يمكن التعبير عنها كالتالى :

التباين والتغاير الكبير لمقدرات الـ OLS:

Large Variances and Covariances of OLS Estimators

للتحقق من أن قيم التباين والتغاير للمقدرات ستكون كبيرة، دعنا نعود للنموذج (1.2.10)، تباين وتغاير $\hat{\beta}_2$ و$\hat{\beta}_3$ معطي كالتالي:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \qquad (12.4.7)$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)} \qquad (15.4.7)$$

$$\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2 \sum x_{3i}^2}} \qquad (17.4.7)$$

حيث r_{23} هو معامل الارتباط بين X_2 و X_3.

يتضح من (12.4.7) و(15.4.7) أن كلما اتجه r_{23} إلى الواحد الصحيح أي أن العلاقة الخطية تزيد، فإن تباين المقدر بين الآتيين يزيد وعندما يؤول r_{23} إلى 1 فإن التباين يؤول إلى ما لا نهاية. وبالتالي فإنه يتضح أيضًا من (17.4.7) أنه كلما زاد r_{23} واقترب من 1 فإن التغاير المحسوب بين المتغيرين يزيد كقيمة مطلقة [لاحظ أن $\operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) \equiv \operatorname{cov}(\hat{\beta}_3, \hat{\beta}_2)$].

السرعة التي يزيد بها التباين والتعاير ممكن حسابها من خلال معامل تضخيم التباين والذي يعرف كالتالي:

$$\mathrm{VIF} = \frac{1}{(1 - r_{23}^2)} \qquad (1.5.10)$$

VIF يوضح المقدار الذي يتضخم به التباين في حالة وجود تعدد في العلاقات الخطية فكلما اقترب r_{23}^2 من 1 فإن VIF يقترب من المالانهاية. أي أنه كلما زاد تعدد العلاقات الخطية كلما زاد تباين المقدر ونهايته تؤول إلى المالانهاية.

كما أنه يمكن بسهولة ملاحظة أنه إذا لم يكن هناك أي ارتباط بين X_2 و X_3 فإن VIF سيساوي 1.

باستخدام هذا التعريف يمكن التعبير عن (12.4.7) و(15.4.7) كالتالي:

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2}\mathrm{VIF} \qquad (2.5.10)$$

$$\operatorname{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2}\mathrm{VIF} \qquad (3.5.10)$$

الذي يوضح أن تباين $\hat{\beta}_2$ و$\hat{\beta}_3$ يناسب مباشرة مع VIF.

لإعطاء مزيد من التوضيح لسرعة زيادة التباين والتغاير مع زيادة r_{23}، دعنا نستعرض جدول (1.10) والذي يعطي بيانات عن هذه التباينات والتغايرات لقيم مختارة لـ r_{23}.

جدول (1.10) أثر زيادة r_{23} على تباين ($\hat{\beta}_2$) والتغايرين ($\hat{\beta}_2$، $\hat{\beta}_3$)

Value of r_{23} (1)	VIF (2)	var ($\hat{\beta}_2$) (3)*	$\frac{\text{var}(\hat{\beta}_2)(r_{23} \neq 0)}{\text{var}(\hat{\beta}_2)(r_{23} = 0)}$ (4)	cov ($\hat{\beta}_2$, $\hat{\beta}_3$) (5)
0.00	1.00	$\frac{\sigma^2}{\sum x_{2i}^2} = A$	—	0
0.50	1.33	1.33 × A	1.33	0.67 × B
0.70	1.96	1.96 × A	1.96	1.37 × B
0.80	2.78	2.78 × A	2.78	2.22 × B
0.90	5.76	5.26 × A	5.26	4.73 × B
0.95	10.26	10.26 × A	10.26	9.74 × B
0.97	16.92	16.92 × A	16.92	16.41 × B
0.99	50.25	50.25 × A	50.25	49.75 × B
0.995	100.00	100.00 × A	100.00	99.50 × B
0.999	500.00	500.00 × A	500.00	499.50 × B

لاحظ أن : $A = \frac{\sigma^2}{\sum x_{2i}^2}$

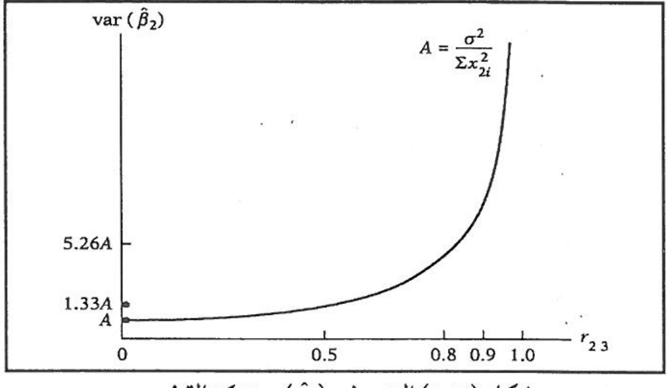
$B = \frac{-\sigma^2}{\sqrt{\sum x_{2i}^2 \sum x_{3i}^3}}$

× = حاصل ضرب

* لايجاد تأثير زيادة r_{23} على تباين $\hat{\beta}_3$ ، لاحظ أن $A = \frac{\sigma^2}{\sum x_{3i}^2}$ عندما يكون r_{23} ولكن معاملات التباين والتغاير تظل كما هي .

كما يتضح من هذا الجدول، زيادة r_{23} له أثر واضح ومحدد على القيم المقدرة للتباينات وتغايرات مقدرات الـ OLS . عندما يكون $r_{23} = 0.5$ ، فإن تباين $\hat{\beta}_2$ يساوي 1.33 مرة قدر التباين عندما يكون $r_{23} = 0$ ، ولكن عندما يصل r_{23} إلى 0.95 فإن التباين يصل إلى عشرة أضعاف قيمته عندما لايكون هناك أي ارتباط .

بالإضافة إلى أنه بزيادة r_{23} من 0.95 إلى 0.995 فإن تقدير التباين يصبح 100 مرة ضعف عندما يكون الارتباط المتعدد = الصفر . نفس التأثير الكبير يحدث أيضًا للتغاير . يمكن ملاحظة كل ذلك في الشكل (2.10) .



شكل (2.10) التغير في Var ($\hat{\beta}_2$) كدالة في r_{23}

النتائج التي حصلنا عليها الآن يمكن بسهولة تطبيقها أيضًا على نموذج به k من المتغيرات. في مثل هذا النموذج، تباين معاملات الـ k متغير، كما سبق تعريفها في (6.5.7)، يمكن التعبير عنها كالتالي:

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\left(\frac{1}{1-R_j^2}\right) \qquad (6.5.7)$$

حيث

$\hat{\beta}_j$ = معامل الانحدار الجزيئي المقدر للمتغير المنحدر X_j

R_j^2 = R^2 في انحدار X_j على باقي المتغيرات المنحدرة، $(k-2)$ متغير

[**لاحظ أن**: هناك $(k-1)$ متغير منحدر في نموذج انحدار يحتوي على k متغير]

$$\sum x_j^2 = \sum(X_j - \bar{X}_j)^2$$

يمكن كتابة (6.5.7) أيضًا كالتالي:

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2}\text{VIF}_j \qquad (4.5.10)$$

يمكن أن نلاحظ من هذه المعادلة الأخيرة، أن Var($\hat{\beta}_j$) يناسب مع σ^2 وVIF ولكن يناسب عكسيًا مع $\sum x_j^2$. وبالتالي، فإن كون تباين $(\hat{\beta}_j)$ كبيراً أو صغيراً فإن ذلك يعتمد على التالي: (1) σ^2، (2) VIF و(3) $\sum x_j^2$ الأخيرة منصوص عليها في الفرض 8 في النموذج التقليدي، عندما ذكرنا أنه كما زاد التباين في المتغيرات المنحدرة كلما قل التباين في معاملات هذه المتغيرات المنحدرة، وذلك بافتراض ثبات العاملين الآخرين، وبالطبع في أن ذلك يزيد من الدقة التي يتم بها تقدير المعامل.

قبل الانتقال إلى نقطة أخرى، جدير بالذكر أن نلاحظ أن مقلوب VIF يطلق عليه معامل التفاوت (TOL) ويحسب كالتالي:

$$\text{TOL}_j = \frac{1}{\text{VIF}_j} = (1 - R_j^2) \qquad (5.5.10)$$

عندما $R_j^2 = 1$ (أي ارتباط تام) فإن $\text{TOL}_j = 0$ وعندما يكون $R_j^2 = 0$ (أي لا يوجد أي ارتباط) فإن JOL_j يساوي 1. ووفقًا لهذه العلاقة بين VIF و TOL فإنه يمكن استخدامها بشكل تبادلي.

فترات ثقة أكثر اتساعاً : Wider Confidence Intervals

فترات الثقة الخاصة بمعالم المجتمع تكون أوسع مع زيادة الأخطاء القياسية، ويمكن ملاحظة ذلك من جدول (2.10). فعلى سبيل المثال، عندما $r_{23} = 0.95$ فإن

فترة الثقة الخاصة بـ β_2 أوسع من نظيرها عندما يكون $r_{23} = 0$ بمعامل يساوي $\sqrt{10.26}$ أي 3 مرات تقريبًا.

وبالتالي في حالة وجود تعدد كبير في العلاقات الخطية ستكون بيانات العينة مؤيدة إيجابيًا لعدد متنوع من الفروض الإحصائية. وبالتالي احتمال قبول فرض خاطئ يزيد (أي يزداد الخطأ من النوع الثاني II).

جدول (2.10) تأثير زيادة العلاقات الخطية على 95% فترة ثقة لـ $\hat{\beta}_2$: $\hat{\beta}_2 \pm 1.96\ \text{Se}(\hat{\beta}_2)$

The Effect of Increasing Collinearity on the 95% CI for β_2 : $\hat{\beta}_2 \pm 1.96\ \text{Se}(\hat{\beta}_2)$

Value of r_{23}	95% confidence interval for β_2
0.00	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.50	$\hat{\beta}_2 \pm 1.96\sqrt{(1.33)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.95	$\hat{\beta}_2 \pm 1.96\sqrt{(10.26)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.995	$\hat{\beta}_2 \pm 1.96\sqrt{(100)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.999	$\hat{\beta}_2 \pm 1.96\sqrt{(500)}\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

لاحظ أننا نستخدم التوزيع الطبيعي، حيث σ^2 بفرض أنها معلومة لحساب التغاير. وبالتالي فإن 1.96 تمثل 95% معامل الثقة للتوزيع الطبيعي.

الأخطاء القياسية المرتبطة بقيم r_{23} المختلفة تم الحصول عليها من الجدول (1.10).

نسب t "غير المعنوية" : "Insignificant" t Ratios

تذكر أنه عند إجراء اختيار فرض، مثلاً، $\beta_2 = 0$، يمكن استخدام النسبة t والتي تساوي $\hat{\beta}_2/\text{se}(\hat{\beta}_2)$ ونقارن هذه القيمة المقدرة لـ t مع القيمة الحرجة من جدول t ولكن كما سبق ورأينا في حالة تعدد العلاقات الخطية، فإن الأخطاء القياسية تزداد بشكل كبير فتكون قيم t صغيرة. وبالتالي سنقبل الفرض العدمي الخاص بأن معلمة المجتمع الحقيقية تساوي الصفر بشكل أكبر[13].

(13) وفقًا لفترات الثقة فإن $\beta_2 = 0$ سيقع داخل منطقة القبول بشكل أكبر كلما زادت درجة تعدد العلاقات الخطية.

قيمة مرتفعة لل R^2 ولكن يصاحبها نسب t المعنوية قليلة :
A High R^2 but few Significant t Ratios

دعنا نعتبر نموذج الانحدار الخطي التالي، والذي يشتمل على k متغير:

$$Y_j = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

في حالة وجود علاقات خطية، من الممكن أن نجد، كما سبق ولاحظنا، أن واحدًا أو أكثر من معاملات الميل الجزيئية ليس له أي معنوية إحصائية منفردًا بناءً على اختيار t.

وعلى الرغم من ذلك، فإننا نجد R^2 في مثل هذه الأحوال مرتفعة، مثلاً، تزيد عن 0.9 وهذا على أساس قيمة اختيار F والتي تجعلنا نرفض الفرض القائل بأن $\beta_2 = \beta_3 = \ldots = \beta_k = 0$. هذه الحالة بالقطع دليل على وجود مشكلة تعدد العلاقات الخطية - والمقصود بالحالة هنا هو قيم t غير معنوية ولكن R^2 مرتفعة وقيم F معنوية.

سنتحدث عن هذه الحالة بالتفصيل في الفقرة التالية، ولكن هذه النتيجة يجب ألا تكون مستغربة خصوصًا بعد المناقشة التي أجريناها من قبل في (الفصل 8) والخاصة بالاختيارات المنفردة والمشتركة. فإذا رجعت إلى تلك النقطة، ستجد أن سبب المشكلة الحقيقية هنا هو التغاير بين المقدرات والذي، وفقًا للمعادلة (17.4.7)، على علاقة بالارتباط بين المتغيرات المنحدرة.

حساسية مقدرات الـ OLS وأخطاؤها القياسية عند حدوث تغير بسيط في البيانات :
Sensitivity of OLS Estimators and Their Standard Errors to small changes in Data

طالما التعدد في العلاقات الخطية غير كامل، فإن تقدير معاملات الانحدار يصبح عملية ممكنة، ولكن التقديرات وأخطاءها القياسية تصبح حساسة جدًا لأى تغيير في البيانات حتى لو كان تغييرًا بسيطًا للغاية.

لتوضيح ذلك، اعتبار جدول (3.10). بناء على هذه البيانات حصلنا على الانحدار المتعدد التالي:

$$\begin{aligned}
\hat{Y}_i &= 1.1939 + 0.4463X_{2i} + 0.0030X_{3i} \\
& \quad\ (0.7737) \quad (0.1848) \quad (0.0851) \\
t &= (1.5431) \quad (2.4151) \quad (0.0358) \\
& R^2 = 0.8101 \qquad r_{23} = 0.5523 \\
& \operatorname{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.00868 \qquad df = 2
\end{aligned} \tag{6.5.10}$$

جدول (3.10)

بيانات افتراضية عن Y ، X_2 ، X_3

Hypothetical data on Y, X_2 and X_3

Y	X_2	X_3
1	2	4
2	0	2
3	4	12
4	6	0
5	8	16

جدول (4.10)

بيانات افتراضية عن Y ، X_2 و X_3

Hypothetical data on Y, X_2, X_3

Y	X_2	X_3
1	2	4
2	0	2
3	4	0
4	6	12
5	8	16

الانحدار الموجود في (6.5.10) يوضح أنه لا توجد معنوية انفرادية لأي من معاملات الانحدار عند مستوى المعنوية 1 أو 5%، على الرغم من أن $\hat{\beta}_2$ لها معنوية عند المستوى 10% بناء على اختيار من طرف واحد فقط للـ t.

والآن دعنا نستعرض جدول (4.10). الفرق الوحيد بين الجدولين (3.10 و 4.10) أن القيمتين الثالثة والرابعة لـ X_3 تم تبديلهما. باستخدام بيانات (4.10) نحصل على التالي:

$$\begin{aligned}
\hat{Y}_i = \; & 1.2108 \; + \; 0.4014X_{2i} + \; 0.0270X_{3i} \\
& (0.7480) \quad (0.2721) \qquad (0.1252) \\
t = \; & (1.6187) \quad (1.4752) \qquad (0.2158) \\
& R^2 = 0.8143 \qquad r_{23} = 0.8285 \\
& \text{cov}(\hat{\beta}_2, \hat{\beta}_3) = -0.0282 \qquad \text{df} = 2
\end{aligned} \tag{7.5.10}$$

وكنتيجة لتغيير بسيط في البيانات، نرى أن $\hat{\beta}_2$ ، والتي كانت معنوية من قبل عند مستوى المعنوية 10% أصبحت غير معنوية الآن حتى عن نفس مستوى المعنوية. أيضًا لاحظ أنه في (6.5.10) كان $\text{Cov}(\hat{\beta}_2, \hat{\beta}_3) = 0.0086$ ولكنه في (7.5.10) يساوي 0.8285. بالمثل الأخطاء القياسية لـ $\hat{\beta}_2$ و $\hat{\beta}_3$.

سبق ولاحظنا أن وجود تعدد كبير في العلاقات الخطية يجعل الفرد غير قادر على تقدير معاملات الانحدار الفردية بشكل دقيق، ولكن التوليفات الخطية المكونة من هذه المعاملات فإنه يمكن تقديرها بشكل أكثر دقة. هذه الحقيقة يمكن توضيحها من الانحدارات (6.5.10) و(7.5.10). ففي الانحدار الأول، مجموع معاملات الميل الجزيئيات هو 0.4493، وفي الثاني هو 0.4284 أي أنهما متساويان عملياً. ليس ذلك

فقط، ولكن أخطاءهما القياسية أيضًا فقد كانت 0.1550 وأصبحت 0.1824 .[14]
وعمومًا لاحظ أيضًا أن معامل X_3 تغير بشكل كبير من 0.003 إلى 0.027 .

عواقب التصغير الجزيئي : Consequences of Micronumerosity

أوضح Goldberger عاقبة مماثلة تمامًا لتعدد العلاقات الخطية، وأطلق عليها اسم التصغير الجزيئي، والمقصود بها مشكلة التحليل بناء على حجم عينة صغير[15]. وننصح القارئ بالرجوع إلى كتاب التحليل الخاص بـ Boldberger ليرى لماذا يساوي Goldberger بين مشكلة تعدد العلاقات الخطية وبين مشكلة التصغير الجزيئي.

6.10 مثال توضيحي: النفقات الاستهلاكية وعلاقتها بالدخل والثروة: AN ILLUSTRATIVE EXAMPLE: CONSUMPTION EXPENDITURE IN RELATION TO INCOME AND WEALTH

لتوضيح النقاط السابق ذكرها، دعنا نرجع مرة أخرى إلى مثال الاستهلاك – الدخل الذي ذكرناه من قبل في (الفصل 3). في جدول (5.10) نستخدم بيانات من جدول (2.3) مع إضافة بيانات عن ثروة المستهلك. إذا افترضنا أن نفقات الاستهلاك مرتبطة خطيًا مع الدخل والثروة، فإنه من جدول (5.10) يمكننا الحصول على الانحدار التالي:

$$\begin{aligned} \hat{Y}_i &= 24.7747 + 0.9415X_{2i} - 0.0424X_{3i} \\ &\quad\ (6.7525) \quad (0.8229) \quad\quad (0.0807) \\ t &= (3.6690) \quad (1.1442) \quad (-0.5261) \\ R^2 &= 0.9635 \quad \bar{R}^2 = 0.9531 \quad \text{df} = 7 \end{aligned} \tag{1.6.10}$$

انحدار (1.6.10) يوضح أن الدخل والثروة معًا يفسران حوالي 96% من التباين في نفقات الاستهلاك على الرغم من أن كلاً من معاملات الميل الخاص بالمتغيرين غير معنوية إحصائيًا بشكل منفرد. والأكثر من ذلك ليس فقط متغير الثروة هو الذي يعتبر

(14) هذه الأخطاء القياسية تم حسابها بناء على:

$$\text{se}(\hat{\beta}_2 + \hat{\beta}_3) = \sqrt{\text{var}(\hat{\beta}_2) + \text{var}(\hat{\beta}_3) + 2\,\text{cov}(\hat{\beta}_2, \hat{\beta}_3)}$$

لاحظ أنه بزيادة الارتباط فإن تباين $\hat{\beta}_2$ و $\hat{\beta}_3$ يزداد أيضًا ولكن هذه التباينات قد تكون متناقضة تمامًا خصوصًا إذا كان لدينا تغاير سلبي بين المعلمين وهذا ما أوضحته النتائج السابقة.

(15) Goldberger, op.cit., pp. 248-250.

متغيرًا غير معنوي إحصائيًا، ولكن لديه أيضًا إشارة خاطئة، فالفرد يتوقع أن يكون لهذا المعامل إشارة موجبة لتوضح العلاقة الطردية بين الاستهلاك والثروة.

جدول (5.10) بيانات افتراضية عن الإنفاق الاستهلاكي Y ، الدخل X_2 والثروة X_3

Hypothetical Data on Consumption Expenditure Y, income X_2 and Wealth X_3

Y, $	X_2, $	X_3, $
70	80	810
65	100	1009
90	120	1273
95	140	1425
110	160	1633
115	180	1876
120	200	2052
140	220	2201
155	240	2435
150	260	2686

جدول (6.10) جدول ANOVA لمثال الاستهلاك - الدخل - الثروة

ANOVA Table for the Consumption - Income - Wealth Example

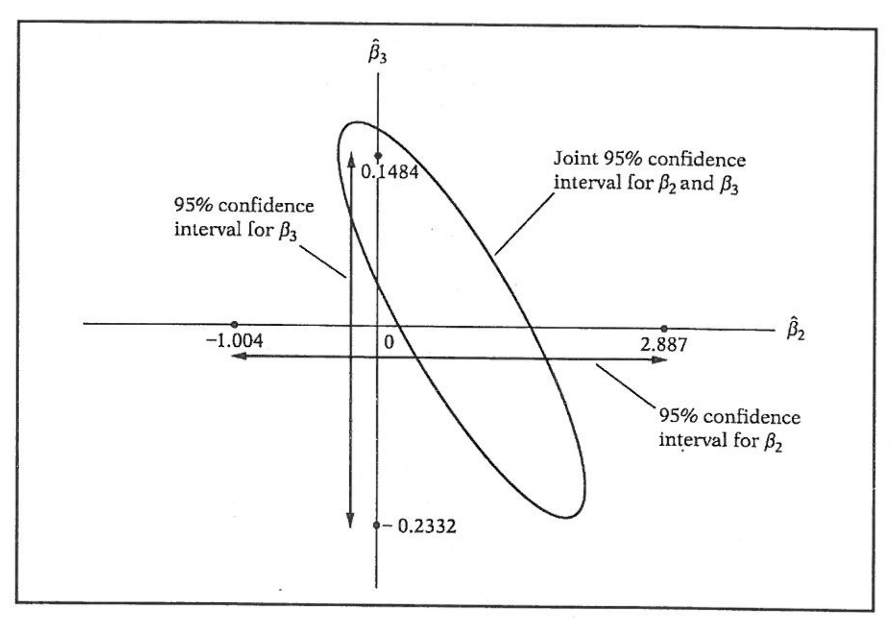
Source of variation	SS	df	MSS
Due to regression	8,565.5541	2	4,282.7770
Due to residual	324.4459	7	46.3494

وعلى الرغم من أن $\hat{\beta}_2$ و $\hat{\beta}_3$ ليس لهما معنوية إحصائية، إلا أننا إذا اختبرنا الفرض $\beta_2 = \beta_3 = 0$ آنيًا، فإننا سنرفض هذا الفرض، كما هو موضح في جدول (6.10). وإن تمت الفروض التقليدية، فإن لدينا

$$F = \frac{4282.7770}{46.3494} = 92.4019 \qquad \textbf{(2.6.10)}$$

قيم F لها معنوية إحصائية مرتفعة.

من الشيق أيضًا استعراض هذه النتائج بيانيًا، انظر شكل (3.10) بناء على الانحدار (1.6.10) فإننا نستطيع تكوين 95% فترة ثقة لكل من β_2 و β_3 باستخدام الأساليب السابق شرحها في (الفصل 8). وكما يتضح من هذه الفترات، فإن كلاً منها تحتوي على الصفر بشكل منفرد. وبالتالي فإننا نقبل الفرض القائل بأن كلاً من معاملات الميل تساوي الصفر بشكل منفرد.



شكل (3.10) فترات الثقة لكل من β_2 و β_3 انفراديًا وفترة الثقة المشتركة (القطع الناقص) لـ β_2 و β_3

Individual CI for β_2 & β_3 and joint CI (ellipse) for β_2 and β_3

ولكن عندما تكون فترة ثقة مشتركة لاختبار الفرض $\beta_2 = \beta_3 = 0$، فإننا لا نستطيع قبول هذا الفرض، حيث إن فترة الثقة المشتركة، والتي تمثل قطعًا ناقصًا، لا تحتوي على نقطة الأصل[16]. وكما سبق أن أوضحنا من قبل، عندما يزداد التعدد في العلاقات الخطية، فإن اختبارات المتغيرات المقدرة الفردية لا يمكن الاعتماد عليها. وفي مثل هذه الحالات، فإن اختبار F الكلي هو الذي سيوضح ما إذا كان المتغير Y على علاقة بالمتغيرات الأخرى أم لا.

مثالنا وضح تمامًا الأثر الكبير الراجع إلى تعدد العلاقات الخطية. فحقيقة أن اختبار F معنوي ولكن قيم t الخاصة بـ X_2 و X_3 غير معنوية كل على حدة، يعني أن هذين المتغيرين مرتبطان بدرجة عالية، بحيث يكون من المستحيل فصل تأثير كل منهما سواء الدخل أو الثروة على الاستهلاك. ففي واقع الأمر إذا قمنا بعمل انحدار لـ X_3 على X_2 سنحصل على:

(16) كما لاحظنا في الفقرة (3.5)، موضوع فترة الثقة المشتركة موضوع متداخل مع النقاط التي نستعرضها في هذا الإطار. القارئ المهتم بذلك عليه العودة إلى المراجع المناسبة.

$$\hat{X}_{3i} = 7.5454 + 10.1909X_{2i}$$
$$(29.4758) \quad (0.1643) \qquad (3.6.10)$$
$$t = (0.2560) \quad (62.0405) \quad R^2 = 0.9979$$

والذي يظهر تعدداً في العلاقات الخطية كامل تقريباً بين X_2 و X_3. دعنا الآن نرى ما الذي سيحدث إذا قمنا بعمل انحدار لـ Y على X_2 بمفردها.

$$\hat{Y}_i = 24.4545 + 0.5091X_{2i}$$
$$(6.4138) \quad (0.0357) \qquad (4.6.10)$$
$$t = (3.8128) \quad (14.2432) \quad R^2 = 0.9621$$

في (1.6.10) متغير الدخل كان معنوياً إحصائياً، والآن أيضاً له معنوية إحصائية عالية، ولكن إذا قمنا بالعكس وحصلنا على انحدار Y على X_3 بدلاً من X_2 سنحصل على:

$$\hat{Y}_i = 24.411 + 0.0498X_{3i}$$
$$(6.874) \quad (0.0037) \qquad (5.6.10)$$
$$t = (3.551) \quad (13.29) \quad R^2 = 0.9567$$

نرى الآن أن الثروة لها تأثير معنوي على نفقات الاستهلاك، في حين أنها في (1.6.10) لم يكن لها أي تأثير على نفقات الاستهلاك.

انحدار (4.6.10) و(5.6.10) يوضحان المشكلة بشكل كبير، فلدينا تعدد كبير في العلاقات الخطية، وإذا أسقطنا أحد هذين المتغيرين المرتبطين سيكون المتغير الآخر له معنوية إحصائية عالية. هذه النتيجة تقترح أن حل مشكلة تعدد العلاقات الخطية قد يكون عن طريق إسقاط أحد هذه المتغيرات المرتبطة من النموذج. ولكننا سنستعرض ذلك بالتفصيل في الفقرة (8.10).

7.10 اكتشاف وجود تعدد في العلاقات الخطية:
DETECTION OF MULTICOLLINEARITY

بعد دراسة طبيعة وعواقب مشكلة تعدد العلاقات الخطية، السؤال المهم والمطروح الآن هو: كيف يمكن للفرد أن يعرف بوجود هذه المشكلة عند الدراسة الفعلية؟ خصوصاً في النماذج التي تحتوي على أكثر من متغيرين اثنين مفسرين؟ دعنا نضع في اعتبارنا النقاط التالية والتي أشار لها Kmenta :

1 - مشكلة تعدد العلاقات الخطية تتوقف على درجة وقوة التعدد وليس في وجود المشكلة في الأصل. فالمعنى الحقيقي لاكتشاف مشكلة تعدد العلاقات الخطية يكمن في تحديد درجة المشكلة وليس فقط وجودها في الأساس.

2 - حيث إن مشكلة تعدد العلاقات الخطية تشير إلى الوضع الذي يفترض فيه أن المتغيرات المفسرة هي متغيرات غير عشوائية فإنها تعتبر صفة أو خاصية للعينة وليس المجتمع.

وبالتالي، فإننا لن نستخدم "اختبار لتعدد العلاقات الخطية" ولكن إذا أردنا الدقة فإننا نقيس درجة التعدد في عينة ما معطاة[17].

وحيث إن تعدد العلاقات الخطية يعتبر ظاهرة من ظواهر العينة، ومع وجود بيانات كثيرة بدون تجارب مصممة من قبل في العديد من العلوم الاجتماعية، فإنه لا يوجد لدينا طريقة وحيدة لاكتشاف أو قياس قوة تعدد العلاقة الخطية. فالذي لدينا هو عدد من القواعد، بعضها محدد بمعادلات وقوانين والبعض الآخر غير محدد، دعنا الآن نستعرض بعضًا من هذه القواعد:

1 - R^2 مرتفعة وعدد قليل من نسب الـ t غير معنوية: High R^2 but few Significant t ratios

كما سبق وذكرنا، فهذه القاعدة تعتبر إشارة "تقليدية" لتعدد العلاقات الخطية. فإذا كانت R^2 كبيرة، مثلاً، تزيد عن 0.8 فإن اختبار F في أغلب الأحوال سيرفض الفرض القائل، فإن معاملات الميل الجزيئية تساوي الصفر آنيًا، ولكن اختبارات t المنفردة ستظهر عدم وجود أو عدد قليل من معاملات الميل المنفردة المعنوية إحصائيًا والمختلفة عن الصفر. هذه الحقيقة تم توضيحها بشكل كبير في مثالنا الخاص بالدخل والاستهلاك والثروة.

على الرغم من منطقية هذه القاعدة، إلا أن عيبها الرئيسي هو أن "هذه القاعدة قوية بالشكل الذي يعتبر تعدد العلاقات الخطية ضارًا ومؤشرًا فقط عندما يكون كل تأثير المتغيرات المفسرة على Y لا يمكن فصله[18].

(17) Jan Kmenta, Elements of Econometrics, 2d ed., Macmillan, New York, 1986, p. 431.

(18) Ibid., p. 439.

2 - ارتباط ثنائي قوي بين المتغيرات المنحدرة؛ High pair-wise correlations among regressors

قاعدة أخرى تقترح أنه إذا كان هناك معامل ارتباط ثنائي أو من الدرجة الصفرية بين متغيرين كثير، مثلاً، يزيد عن 0.8 فإن مشكلة تعدد العلاقات الخطية يعتبر مشكلة خطيرة. المشكلة في هذه القاعدة أنه على الرغم من أن وجود ارتباط من الدرجة الصفرية عال يعتبر مؤشراً لوجود تعدد في العلاقات الخطية، إلا أن ذلك ليس بالضرورة صحيحاً في كل الأحوال. للتعبير بشكل علمي أكثر عن هذا الأمر، فإن الارتباط الصغرى يعتبر شرطاً كافياً ولكن غير ضروري لوجود تعدد في العلاقات الخطية، لأنه من الممكن أن يكون هناك تعدد في العلاقات الخطية حتى إذا كان معامل الارتباط البسيط أو صفري الدرجة صغيراً (أقل من 0.5) لتوضيح ذلك دعنا نفترض أن لدينا نموذجاً يشتمل على أربعة متغيرات كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

وافترض أن:

$$X_{4i} = \lambda_2 X_{2i} + \lambda_3 X_{3i}$$

مع افتراض أن λ_2 و λ_3 ثوابت لا يتساويان مع الصفر. وبما أن X_4 هي توليفة خطية من X_2 و X_3 فإن $R^2_{4.23} = 1$، وذلك يمثل معامل تحديد انحدار X_4 على كل من X_2 و X_3.

والآن بالعودة إلى المعادلة (5.11.7) من (الفصل 7)، فإنه يمكن كتابتها كالتالي:

$$R^2_{4.23} = \frac{r^2_{42} + r^2_{43} - 2r_{42}r_{43}r_{23}}{1 - r^2_{23}} \qquad (1.7.10)$$

ولكن بما أن $R^2_{4.23} = 1$ حيث يوجد تعدد كامل في العلاقات الخطية، فإننا نحصل على:

$$1 = \frac{r^2_{42} + r^2_{43} - 2r_{42}r_{43}r_{23}}{1 - r^2_{23}} \qquad (2.7.10)$$

ليس من الصعب أن نرى (2.7.10) تتحقق أيضاً باستخدام $r_{42} = 0.5$ و $r_{43} = 0.5$ و $r_{23} = -0.5$ والتي لا تعتبر قيماً مرتفعة.

وبالتالي، فإنه في النماذج التي تشتمل على أكثر من متغيرين اثنين، فإن الارتباط الصفري أو البسيط لن يكون دليلاً كافياً لوجود تعدد العلاقات الخطية. وبالطبع إذا كان هناك متغيران مفسران اثنان فقط، فإن الارتباط الصفري سيكون كافياً.

3 - اختبارات الارتباطات الجزيئية : Examination of partial correlations

حيث إن هناك مشكلة في الارتباط الصفرية والسابق ذكرها لاحقًا، فإن Glauber وFarrar اقترحا ضرورة الاعتماد على معاملات الارتباط الجزيئية[19]. وبالتالي في انحدار y على كل من X_2، X_3 و X_4، فإن وجود $R^2_{1.234}$ مرتفع مع انخفاض في قيم $r^2_{12.34}$ و $r^2_{13.24}$ و $r^2_{14.23}$ قد يعطي إشارة أن هذه المتغيرات X_2 وX_3 وX_4 مرتبطة بشكل كبير، وعلى الأقل واحد من هذه المتغيرات زائد عن الحاجة.

وعلى الرغم من أن دراسة الارتباطات الجزيئية قد تكون مساعدة كدليل على وجود تعدد في العلاقات الخطية ولكن ما الذي سيحدث إذا كان كل من R^2 والارتباطات الجزيئية عالية أو مرتفعة القيمة. وأيضًا C. Robert Wichers أوضح[20] أن اختيار معاملات الارتباط الجزيئية متفاعل مع أنماط مختلفة من التعدد في العلاقات الخطية.

اختيار Farrar-Glauber وجد له العديد من الانتقادات في دراسات T. Krishna Kumar[21] و John O'Hagan و Brendan McCabe[22].

4 - الانحدارات المساعدة : Auxiliary Regressions

بما أن تعدد العلاقات الخطية يظهر نتيجة وجود أكثر من متغير منحدر في صورة توليفة خطية قوية أو تامة من المتغيرات المنحدرة الأخرى، فإنه أحد الطرق المستخدمة لمعرفة هذا المتغير وتحديده هي القيام بانحدار لكل متغير X على باقي المتغيرات المنحدرة وحساب R^2 الخاصة بكل انحدار، والتي سنعرفها بـ R_i^2، وسيطلق على كل انحدار من هذه الانحدارات مصطلح الانحدار المساعد، أي المساعد للانحدار الرئيسي لـ Y على الـ X's. ووفقًا للعلاقة بين F، R^2 المحددة في (11.5.8)، المتغير

$$F_i = \frac{R^2_{x_i \cdot x_2 x_3 \cdots x_k}/(k-2)}{\left(1 - R^2_{x_i \cdot x_2 x_3 \cdots x_k}\right)/(n-k+1)} \qquad (3.7.10)$$

(19) D. E. Farrar and R. R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," Review of Economics and Statistics, vol. 49, 1967, pp. 92-107.

(20) "The Detection of Multicollinearity: A Comment," Review of Economics and Statistics, vol. 57, 1975, pp. 365-366.

(21) "Multicollinearity in Regression Analysis," Review of Economics and Statistics, vol. 57, 1975, pp. 366-368.

(22) "Tests for the Severity of Multicollinearity in Regression Analysis: A Comment," Review of Economics and Statistics, vol. 57, 1975, pp. 368-370.

يتبع توزيع F بدرجات حرية $k-2$ و $n-k+1$. في المعادلة (3.7.10) n عبارة عن حجم العينة ، k عدد المتغيرات المفسرة مضافًا إليه الجزء الثابت المقطوع من المحور الصادي و $R^2_{x_i, x_2, x_3 \ldots x_k}$ هو معامل التحديد في انحدار المتغير X_i على باقي المتغيرات X .[23]

إذا كانت قيمة F المحسوبة تزيد عن القيمة F_i الحرجة عند مستوى المعنوية المحدد، فإن ذلك يعني أن تلك X_i مرتبطة مع باقي الـ X's إذا لم تزد عن القيمة الحرجة F_i ، فإننا نستنتج أن X_i غير مرتبطة مع باقي الـ X's، وبالتالي يحتفظ بالمتغير X في النموذج.

إذا كانت F_i معنوية إحصائيًا ، فإننا مازلنا نحتاج أن نقرر ما إذا كان المتغير X_i يظل في النموذج أم نسقطه من النموذج. هذا السؤال سيتم تناوله بشكل مفصل في الفقرة (8.10).

ولكن هذه الطريقة أيضًا لها عيوبها كالتالي :

إذا كان تعدد العلاقات الخطية يتضمن عددًا محدودًا من المتغيرات ، فإن الانحدارات المساعدة لن تعاني من هذه المشكلة ، والمعاملات المقدرة ستظهر طبيعة الارتباط الخطي بين المتغيرات المنحدرة. ولكن للأسف إذا كانت هناك أشكال أكثر تعقيدًا من العلاقات الخطية ، فإن تلك الطريقة قد لا تثبت شيئًا وقد يكون صعب استخدامها لتحديد العلاقات التداخلية المنفصلة[24].

بدلاً من استخدام جميع قيم R^2 المساعدة السابقة ، من الممكن الاعتماد على قاعدة **klien** والتي تقترح أن تعدد العلاقات الخطية قد يسبب مشكلة حقيقية في الدراسة إذا كانت R^2 التي نحصل عليها من الانحدار المساعد أكبر من R^2 الكلية التي نحصل عليها من انحدار Y على كل المتغيرات المنحدرة[25]. وبالطبع مثل باقي القواعد المذكورة سابقًا يجب أن يتم التعامل مع هذه القاعدة أيضًا بحذر شديد.

---

(23) على سبيل المثال ، $R^2_{x_2}$ يمكن الحصول عليها من عمل انحدار X_{2i} كالتالي :

$$X_{2i} = a_1 + a_3X_{3i} + a_4X_{4i} + \ldots + a_kX_{ki} + \hat{u}_i$$

(24) George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lüthepohl, and Tsoung-Chao Lee, Introduction to the Theory and Practice of Econometrics, John Wiley & Sons, New York, 1982, p. 621.

(25) Lawrence R. Klien, An Introduction to Econometrics, Prentice-Hall, Englewood Cliffs, N.J., 1962, p. 101.

5 - القيم المميزة ومؤشر الشرط : Eigenvalues and condition index

إذا نظرت بدقة لمخرجات برنامج SAS والخاص بدالة إنتاج Cobb-Douglas والمعطاة في الملحق 5A.7 سترى أن الـ SAS يستخدم القيم المميزة ومؤشر الشرط للتعرف على مشكلة تعدد العلاقات الخطية. لن نناقش هنا القيم المميزة، حيث إننا سنضطر إلى دراسة جبر المصفوفات، وهو الموضوع خارج إطار الكتاب الحالي. وعمومًا من هذه القيم المميزة نستطيع اشتقاق قيم k معروفة باسم الرقم الشرطي والمعرف كالتالي:

$$k = \frac{\text{أكبر قيمة مميزة}}{\text{أصغر قيمة مميزة}}$$

ومؤشر الشرط (CI) يعرف كالتالي:

$$CI = \sqrt{\frac{\text{أكبر قيمة مميزة}}{\text{أصغر قيمة مميزة}}} = \sqrt{k}$$

إذن هذه قاعدة أخرى يمكن أن نستخدمها، إذا كانت k بين 100 و 1000، فإن لدينا تعددًا في العلاقات الخطية يتراوح من المتوسط والقوى، وإذا زادت عن 1000 فإن لدينا تعددًا قويًا جدًا في العلاقات الخطية وبالمثل أيضًا إذا كان CI($=\sqrt{k}$) من 10 و 30 فإن لدينا تعددًا في العلاقات الخطية يتراوح بين المتوسط والقوى وإذا زاد عن 30 فإن لدينا تعددًا قويًا في العلاقات الخطية.

وكمثال توضيحي، افترض أن $k = \frac{3}{0,00002422}$ أو تساوي 123,864 و $CI = \sqrt{123,864}$ أي يساوي 352، نلاحظ أن كلاً من k و CI يمكن حسابهما بناء على أكبر قيمة من القيم المميزة وأي قيمة مميزة أخرى، وذلك واضح في مخرجات البرنامج (لاحظ أن: في مخرجات البرنامج لا يوجد حساب صريح للـ k ولكن يمكن حسابها بسهولة عن طريق تربيع الـ CI) ولاحظ أن القيمة المميزة الأصغر (بالنسبة لعلاقتها مع أكبر قيمة مميزة) يعتبر مؤشرًا عامًا على الاقتراب من الارتباط الخطي في البيانات.

بعض الكتاب يعتقدون أن مؤشر الشرط هو أفضل الطرق المتاحة للتعرف على تعدد العلاقات الخطية. ولكن هذا الرأى غير منتشر في أوساط عديدة بالنسبة لنا، فإن CI هو مجرد قاعدة للتحديد وإن كان من أكثرها تعقيدًا.

ولمزيد من التفاصيل، يمكن أن يقرأ الباحث في المراجع المذكورة في هذا الشأن[26].

(26) انظر في D. A. Belsley, E. Kuh, and R. E. Welsch, Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley & Sons, New York, 1980, Chap. 3 مع ملاحظة ان هذا الكتاب ليس للمبتدئين.

6 - التفاوت ومعامل تضخم التباين Tolerance and variance Inflation Factor

سبق وأن استعرضنا سريعًا TOL و VIF . فكلما زاد R_j^2 واقترب من الواحد الصحيح، حيث R_j^2 هو معامل التحديد في الانحدار الخاص بالمتغير المنحدر X_j على باقي المتغيرات المنحدرة الموجودة في النموذج، فإن الارتباط بين X_j وباقي المتغيرات المنحدرة يزداد ويزداد أيضًا VIF ونهاية تؤول إلى اللا نهاية.

بعض الكتاب يستخدمون VIF كمؤشر على وجود تعدد في العلاقات الخطية، فكلما زادت قيمة VIF_j كلما كان ذلك دليلاً على وجود مشاكل من ارتباط المتغير X_j.

وكقاعدة عملية، فإنه إذا زادت قيمة VIF للمتغير عن 10، وذلك سيحدث متلازمًا مع R_j^2 أكبر من 0.9، فإن المتغير يقال أنه مرتبط خطيًا بدرجة كبيرة مع باقي المتغيرات[27].

بالطبع من الممكن استخدام TOL_j كمقياس لتعدد العلاقات الخطية من خلال علاقته بـ VIF_j، فكلما اقترب TOL_j من الصفر كلما زادت درجة تعدد العلاقة الخطية بين هذا المتغير وبقي المتغيرات الأخرى. على الجانب الآخر، كلما اقترب TOL_j من الواحد الصحيح، كلما زاد دليل أن X_j ليس مرتبطًا خطيًا مع باقي المتغيرات المنحدرة.

VIF (أو التفاوت) كمقياس للتعدد في العلاقات الخطية ليس خاليًا من العيوب أو المشاكل، فكما يتضح من (4.5.10) فإن Var($\hat{\beta}_j$) يعتمد على ثلاثة عوامل: σ^2، $\sum x_j^2$ و VIF_j، وبالتالي، فإنه إذا كانت قيمة VIF كبيرة، فإنها قد تتوازن مع σ^2 صغيرة أو $\sum x_j^2$ كبيرة. بمعنى آخر، القيمة المرتفعة لـ VIF ليست شرطًا كافيًا ولا ضروريًا للحصول على تباين مرتفع وأخطاء قياسية مرتفعة. وبالتالي فإن التعدد الكبير في العلاقات الخطية والذي نتعرف عليه من خلال قيمة مرتفعة للـ VIF ليس بالضرورة يؤدي إلى أخطاء قياسية مرتفعة. في كل الأحوال، المصطلحات "مرتفع" و "منخفض" لابد أن تستخدم بشكل نسبي.

لتلخيص كل ما سبق مناقشته عن اكتشاف التعدد في العلاقات الخطية، نحب أن نؤكد على أن هناك طرقًا عديدة تم مناقشتها تعتبر ضرورية للتعرف على ظاهرة تعدد العلاقات الخطية، ولكننا لا نستطيع تحديد أي من هذه الطرق سيصلح أفضل

(27) انظر في David G. Kleinbaum, Lawrence L. Kupper, and Keith E. Muller, Applied Regression Analysis and other Multivariate Methods, 2d ed., PWS-Kent, Boston, Mass., 1988, p. 210.

وفقًا لدراسة ما معطاة. بالإضافة إلى أنه لا يوجد حل عملي حتى الآن للتعرف على مشكلة وجود تعدد في العلاقات الخطية في بيانات عملية ليست ناتجة عن تجربة إحصائية، وبالتالي لا تخضع لسيطرة الباحث. وتعتبر المشكلة الأخيرة ليست ناتجة عن تجربة إحصائية، وبالتالي فإنها من أهم مشاكل البحث في العلوم الاجتماعية.

ومرة أخرى وبشكل تجميعي، فإن Goldberger أوضح عددًا من الطرق لاكتشاف وجود التصغير الجزيئي، مثل تحديد القيم الحرجة للعينة من الحجم n^*، وتكون مشكلة التصغير الجزيئية موجودة إذا كان حجم القيمة الأصلي n أصغر من n^*. وبالتالي فالنقطة التي يشير إليها Goldberger هي أن حجم القيمة الصغيرة وعدم وجود تنوع في المتغيرات المفسرة قد يؤدي إلى مشاكل لا تقل أهمية وخطورة عن المشاكل الناتجة من التعدد في العلاقات الخطية.

8.10 إجراءات علاجية : REMEDIAL MEASURES

ما الذي يمكن فعله إذا وجد لدينا تعدد كبير في العلاقات الخطية؟ هناك اختياران: (1) عدم فعل أي شئ أو (2) اتباع بعض القواعد العامة.

عدم فعل أي شئ : Do Nothing

مدرسة "عدم فعل أي شئ" تتضح أفكارها من خلال كتابات Blanchand كالتالي[28]:

عندما يقوم الطلاب بإجراء أول انحدار لهم بالمربعات الصغري العادية، أولى المشاكل التي تواجههم هي تعدد العلاقات الخطية. فالعديد منهم يستنتج أن هناك خطأ ما في الـ OLS، والبعض منهم يحاول البحث عن طرق فعالة للتغلب على هذه المشكلة. ولكن نقول لهم هذا خطأ، فالتعدد في العلاقات الخطية "قدر" من عند الخالق وليست مشكلة في الـ OLS أو أي أسلوب إحصائي آخر.

والذي يقصده Blanchard هنا هو أن تعدد العلاقات الخطية هو مشكلة عادية للبيانات (التصغير الجزيئي مرة أخرى) وأحيانًا يكون ليس لدينا خيار آخر للتعامل مع البيانات الخاصة بالتحليل العملي.

(28) Blanchard, O. J., Comment, Journal of Business and Economic Statistics, vol. 5, 1967, pp. 449–451. The quote is reproduced from Peter Kennedy, A Guide to Econometrics, 4th ed., MIT Press, Cambridge, Mass., 1998, p. 190.

أيضًا الموضوع لن يتوقف بوجود عدد من معاملات انحدار النموذج غير المعنوية إحصائيًا أو حتى إذا لم نستطع تقدير واحد أو أكثر من معاملات الانحدار بدقة عالية، ففي النهاية نستطيع تقدير. توليفة خطية (أي دالة قابلة للتقدير) من المتغيرات بكفاءة نسبية عالية. كما رأينا في (3.2.10) فإننا نستطيع تقدير α حتى ولم نستطع تقدير مكوناتها بشكل دقيق ومنفصل. وأحيانًا يكون ذلك هو أفضل ما يمكن عمله وفقًا للبيانات المتاحة[29].

طرق قاعدة الإبهام : Rule-of-Thumb Procedures

من الممكن للباحث أن يستخدم الطرق التالية لتحديد شكل تعدد العلاقات الخطية، والنجاح في استخدام القاعدة يعود إلى درجة قوة مشكلة تعدد العلاقات الخطية.

1 - معلومات مسبقة : A priori information

افترض أن لدينا النموذج التالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

حيث Y = الاستهلاك، X_2 = الدخل و X_3 = الثروة. كما لاحظنا من قبل، فإن المتغيرين الدخل والثروة بينهما ارتباط قوي. ولكن افترض أن لدينا معلومات مسبقة بأن $\beta_3 = 0.10\beta_2$ أي أن معدل التغير في الاستهلاك بالنسبة للثروة يعادل عشر معدله بالنسبة للدخل. **وبالتالي يمكننا الآن القيام بالانحدار التالي :**

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + 0.10\beta_2 X_{3i} + u_i \\ &= \beta_1 + \beta_2 X_i + u_i \end{aligned}$$

حيث $X_i = X_{2i} + 0.1X_{3i}$. وبمجرد أن نحصل على $\hat{\beta}_2$ فإنه يمكننا تقدير $\hat{\beta}_3$ من العلاقة المسبقة الموجودة بين β_2 و β_3.

كيف يمكننا الحصول على المعلومة المسبقة؟ يمكن أن نحصل عليها من دراسات سابقة، والتي لم تكن فيها مشكلة الارتباط المتعدد موجودة بها بدرجة كبيرة أو من النظرية المرتبطة بالمجال الخاص بالدراسة العملية. فمثلاً في دالة إنتاج Cobb-Douglas

(29) لمناقشة مثمرة انظر في: Conlisk, J., "When Collinearity is Desirable," Western Economic Journal, vol. 9, 1971, pp. 393–407.

(1.9.7) إذا توقع الفرد ثبات العائد، وبالتالي $(\beta_2 + \beta_3) = 1$ فإنه من الممكن عمل (14.7.8) بانحدار نسبة الناتج إلى العمالة على نسبة رأس المال إلى العمالة. إذا وجد تعددًا في العلاقات الخطية بين العمالة ورأس المال، كما هو الحال في العديد من بيانات العينة، فمثل هذه التحويلة قد تقلل أو تنهى مشكلة الارتباط المتعدد. ولكن لابد من التنويه هنا عن المشاكل المترتبة على فرض مثل هذه القيود المسبقة، "... بما أننا في العموم نرغب في اختبار النظرية الاقتصادية وعواملها التنبؤية بدلاً من فرض هذه النظرية على البيانات، وهذه النظرية قد لا تكون بالضرورة صحيحة"[30].

عمومًا، تعرفنا في الفقرة 7.8 على كيفية التحقق من صحة هذه القيود.

2 - المزج بين البيانات المقطعية وبيانات السلاسل الزمنية :

Combining cross-sectional and time series Data

كبديل لأسلوب المعلومات المسبقة السابق ذكره، ممكن المزج بين البيانات المقطعية وبيانات السلاسل الزمنية، وهذا يسمى بتجميع البيانات أو البيانات المجمعة. افترض أننا نريد دراسة الطلب على السيارة في الولايات المتحدة الأمريكية، وافترض أن لدينا بيانات سلسلة زمنية على عدد السيارات المباعة، متوسط سعر السيارة ودخل المستهلك. وافترض أيضًا التالي:

$$\ln Y_t = \beta_1 + \beta_2 \ln P_t + \beta_3 \ln I_t + u_t$$

حيث Y = عدد السيارات المباعة، P = متوسط السعر، I = الدخل و t = الزمن. هدفنا هو تقدير مرونة السعر β_2 ومرونة الدخل β_3.

في بيانات السلسلة الزمنية متغيرات السعر والدخل يوجد بينهما ارتباط قوي. وبالتالي، إذا قمنا بإجراء الانحدار السابق ستواجهنا المشكلة المعتادة والخاصة بتعدد العلاقات الخطية. وكطريقة للتخلص من هذه المشكلة والتي اقترحها Tobin.[31] أنه إذا كان لدينا بيانات مقطعية (مثلاً بيانات مجمعة من مؤسسات استهلاكية أو دراسات ميزانية سواء خاصة بهيئات حكومية أو خاصة)، فإنه يمكن استخدامها

(30) Mark B. Stewart and Kenneth F. Wallis, Introductory Econometrics, 2d ed., John Wiley & Sons, A Halstead Press Book, New York, 1981, 9. 154.

(31) J. Tobin, "A Statistical Demand Function for Food in the U.S.A.," Journal of the Royal Statistical Society, Ser. A, 1950, pp. 113–141.

لتقدير مرونة الدخل β_2، حيث إن في هذه البيانات المأخوذة عند نقطة زمنية واحدة فقط، الأسعار لن تتغير كثيراً. دعنا نفترض أن تقدير البيانات المقطعية لمرونة الدخل هو $\hat{\beta}_3$. باستخدام هذا التقدير، يمكننا كتابة انحدار السلاسل الزمنية السابق كالتالي:

$$Y_t^* = \beta_1 + \beta_2 \ln P_t + u_t$$

حيث $Y^* = \ln Y - \hat{\beta}_3 \ln I$، أي أن Y^* تمثل قيمة Y بعد تخليصها من أثر الدخل. تستطيع الآن الحصول على مقدر لمرونة السعر β_2 من الانحدار السابق.

على الرغم من سهولة الأسلوب السابق، إلا أن تجميع بيانات السلاسل الزمنية والبيانات المقطعية بالأسلوب السابق ذكره، قد يؤدي إلى وجود مشكلة في تفسير النتائج، حيث إنه يفترض أن مقدار مرونة الدخل باستخدام البيانات المقطعية هو نفسه المقدار الذي ستحصل عليه إذا استخدمنا تحليلاً يقتصر على السلاسل الزمنية[32]. ولكن عموماً وبغض النظر عن تلك الملاحظة، فإن هذا الأسلوب تم استخدامه في العديد من التطبيقات، ومن المهم التطرق إليه خصوصاً في الحالات التي لا تتغير فيها مقدرات البيانات المقطعية من قطاع إلى قطاع آخر. كمثال لهذا الأسلوب، انظر في تمرين 26.10.

3 - إسقاط متغير أو أكثر وتحيز التوصيف: Dropping a variable(s) and specification bias

عندما يوجد تعدد قوي في العلاقات الخطية، إحدى الطرق "البسيطة" هي إسقاط واحد من المتغيرات المرتبطة. فمثلاً في مثالنا التوضيحي الخاص بالدخل - الاستهلاك - الثروة، عندما نسقط متغير الثروة من التحليل، ونحصل على انحدار (4.6.10) نجد أن متغير الدخل معنوي إحصائيًا بدرجة عالية كما كان الحال في النموذج الأصلي. ولكن مع إسقاط متغير من النموذج.

ممكن أن نقع في تحيز التوصيف أو خطأ التوصيف. فتحيز التوصيف يأتي من التوصيف غير السليم للنموذج المستخدم في التحليل.

وبالتالي، إذا نصت النظرية الاقتصادية بأن الدخل والثروة يجب أن يستخدما معًا في النموذج الذي يفسر نفقات الاستهلاك، فإن إسقاط متغير الثروة من النموذج

(32) لقراءة المزيد عن الأسلوب التجميعي وتطبيقاته، انظر
Edwin Kuh, Capital Stock Growth: A Micro-Econometric Approach, North-Holland Publishing Company, Amsterdam, 1963, Chaps. 5 and 6.

سيؤدي إلى وجود تحيز في التوصيف، على الرغم من أننا سنناقش بالتفصيل خطأ التوصيف في (الفصل 13). فإننا استعرضناه سريعًا في الفقرة 7.7. فمثلاً إذا كان النموذج الصحيح هو:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

لكن بالخطأ فإننا نقدر النموذج

$$Y_i = b_1 + b_{12} X_{2i} + \hat{u}_i \quad (1.8.10)$$

فإنه يمكن ملاحظة أن (انظر ملحق 1A.13)

$$E(b_{12}) = \beta_2 + \beta_3 b_{32} \quad (2.8.10)$$

حيث b_{32} = ميل معامل انحدار X_3 على X_2. وبالتالي يتضح من (2.8.10) أن b_{12} سيكون مقدرًا متحيزًا لـ β_2 طالما b_{32} ستكون مختلفة عن الصفر (إذا كان مفروض أن β_3 لاتساوي الصفر، لأنه بخلاف ذلك لا يوجد أي معنى من إدخال X_3 في النموذج)[33]. بالطبع إذا كان b_{32} يساوي الصفر فإنه لا يوجد لدينا مشكلة تعدد في العلاقات الخطية. ويتضح أيضًا من (2.8.10) أنه إذا كان كل من b_{32} و β_3 موجبين (أو كلاهما سالبان) فإن $E(b_{12})$ سيكون أكبر من β_2، وبالتالي في المتوسط b_{12} ستكون تقدير زائد دائمًا عن القيمة الحقيقية لـ β_2 مما يؤدي إلى وجود تحيز موجب. بالمثل إذا كان حاصل ضرب $b_{32}\beta_3$ سالبًا فإنه في المتوسط b_{12} سيكون تقديره أقل دائمًا من القيمة الحقيقية لـ β_2 مما يؤدي إلى وجود تحيز سالب.

من المناقشة السابقة يتضح أن إسقاط متغير من النموذج كوسيلة للتغلب على مشكلة تعدد العلاقات الخطية قد يؤدي إلى خطأ في التوصيف، وهنا قد يكون ذلك أكثر خطورة من المشكلة الأصلية، فيكون العلاج أكثر خطورة من المرض نفسه حيث انه إذا كان تعدد العلاقات الخطية يقلل من دقة تقدير معالم النموذج، فإن حذف متغير من النموذج قد يؤدي إلى الوصول إلى قيم غير صحيحة تمامًا عن معالم المجتمع. تذكر أن مقدرات OLS ستظل BLUE حتى في وجود علاقات خطية متعددة.

4 - تحويل المتغيرات: Transformation of Variables

افترض أن لدينا بيانات سلاسل زمنية على النفقات الاستهلاكية والدخل والثروة. أحد أسباب وجود تعدد في العلاقات الخطية بين الدخل والثروة في مثل

(33) لاحظ أيضًا أنه اذا كانت b_{32} تتقرب من الصفر مع زيادة حجم العينة فإن b_{12} لن يكون فقط مقدر متحيز ولكن سيكون مقدر غير متسق أيضًا.

هذا النوع من البيانات هو أنه بمرور الزمن، فإن كلاً من المتغيرين يميلان للتحرك في نفس الاتجاه. دعنا نستعرض أحد الأساليب التي يمكن اتباعها لتقليل التبعية كالتالي :

إذا كانت العلاقة على الشكل

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad (3.8.10)$$

فإنها تتحقق عند الزمن t وعند الزمن $t-1$ أيضًا حيث إن نقطة الأصل t هي نقطة اختيارية. وبالتالي فإن لدينا :

$$Y_{t-1} = \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \qquad (4.8.10)$$

إذا قمنا بطرح (4.8.10) من (3.8.10) سنحصل على :

$$Y_t - Y_{t-1} = \beta_2(X_{2t} - X_{2,t-1}) + \beta_3(X_{3t} - X_{3,t-1}) + v_t \qquad (5.8.10)$$

حيث $v_t = u_t - u_{t-1}$، لمعادلة (5.8.10) معروفة باسم معادلة الفروق الأولى، حيث إننا نقوم بعمل الانحدار على الفروق بين مشاهدات المتغير وليس القيمة الأصلية للمتغير.

نموذج انحدار الفروق الأولى يقلل دائمًا من حدة مشكلة تعدد العلاقات الخطية، حيث إنه على الرغم من أن قيم X_2 و X_3 قد تكون مرتبطة بدرجة عالية، فإنه لا يوجد أي سبب مسبق يجعل فروقهم أيضًا مرتبطة بدرجة عالية.

وكما سنرى في الفصول الخاصة بالاقتصاد القياسي الخاص بالسلاسل الزمنية، فإنه من أهم مميزات تحويل الفروق الأولى هي تحويل السلسلة الزمنية من سلسلة غير ساكنة إلى أخرى ساكنة. في هذه الفصول، سنرى الأهمية الخاصة بكون السلسلة الزمنية. كما رأينا في (الفصل 1) وبشكل عام، فإن السلسلة الزمنية، على سبيل المثال Y_t، يقال عنها إنها سلسلة ساكنة إذا كان الوسط الحسابي والتباين لا يتغيران بشكل منتظم عبر الزمن . . .

من التحويلة الأخرى التي تستخدم كثيرًا من التطبيق العلمي تحويلة النسبة، دعنا نعتبر النموذج التالي :

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad (6.8.10)$$

حيث Y نفقات الاستهلاك بالدولار، X_2 هو GDP و X_3 عدد السكان الإجمالي. بما أن GDP وعدد السكان يزايد بمرور الزمن، فإنه من الأرجح أن يكونا مرتبطين. أحد "حلول" هذه المشكلة هو أن يتم التعبير عن النموذج بالنسبة للفرد الواحد، أي يتم قسمة (6.8.10) على X_3 فنحصل على التالي:

$$\frac{Y_t}{X_{3t}} = \beta_1\left(\frac{1}{X_{3t}}\right) + \beta_2\left(\frac{X_{2t}}{X_{3t}}\right) + \beta_3 + \left(\frac{u_t}{X_{3t}}\right) \qquad (7.8.10)$$

مثل هذه التحويلة قد تؤدي إلى تقليل الارتباط في المتغيرات الأصلية. ولكن سواء تحويلة الفروق الأولى أو النسبة ليستا خاليتين من المشاكل. فمثلاً، مقدار الخطأ v_t والموجود في (5.8.10) قد لا يكون مستوفيًا للشروط الخاصة بنموذج الانحدار الخطي التقليدي، وبالأخص أن يكون الخطأ u_t غير مرتبط ذاتيًا. وكما سنرى في (الفصل 12)، إذا كان حد الخطأ u_t غير مرتبط ذاتيًا فإن مقدار الخطأ v_t الذي حصلنا عليه سابقًا سيكون مرتبطًا ذاتيًا في أغلب الأحيان. وبالتالي مرة أخرى، يصبح العلاج أكثر خطورة من المرض نفسه. والأكثر من ذلك هو أن هناك فقدًا لمشاهدة واحدة نتيجة استخدام الفروق الأولى، وبالتالي فإن درجات الحرية تقل بواحد صحيح.

في **العينات الصغيرة**، يجب أن يضع الباحث هذا الأمر في الاعتبار. والأكثر من ذلك أن تحويلة الفروق الأولى قد لا يجوز استخدامها في حالة البيانات المقطعية، حيث لا يوجد ترتيب منطقي للمشاهدات.

بالمثل في نموذج النسبة (7.8.10)، مقدار الخطأ.

$$\left(\frac{u_t}{X_{3t}}\right)$$

سيكون غير ثابت في التباين حتى إذا كان مقدار الخطأ الأصلي u_t ثابتًا في التباين. وسنرى ذلك تفصيلياً في (الفصل 11). مرة أخرى فعلاج مشكلة تعدد العلاقات الخطية قد تكون أسوأ من المشكلة نفسها. باختصار يجب على الباحث أن يكون حريصًا عندما يستخدم طريقة الفروق الأولى أو طريقة النسبة لتحويل البيانات لحل مشكلة تعدد العلاقات الخطية.

5 – بيانات جديدة أو إضافية : Additional or new data

بما أن تعدد العلاقات الخطية يعتبر ظاهرة من ظواهر العينة، فإنه من الممكن في عينة أخرى بها نفس المتغيرات يكون الارتباط غير قوى كما في العينة الأولى. ففي

بعض الأحيان، يكون من السهل زيادة حجم العينة (إذا كان ذلك ممكنًا) مما يقلل من مشكلة الارتباط المتعدد.

فعلى سبيل المثال، في النموذج ثلاثي المتغيرات رأينا التالي:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{23}^2)}$$

والآن وبزيادة حجم العينة، Σx_{2i}^2 سيزداد أيضًا (لماذا؟)

وبالتالي، فإنه بالنسبة لقيمة معطاة لك r_{23}، تباين $\hat{\beta}_2$ سيقل مما يقلل الخطأ القياسي، مما سيجعلنا قادرين على تقدير β_2 بشكل أكثر دقة.

للتوضيح، دعنا نعتبر نموذج الانحدار التالي والخاص بنفقات الاستهلاك y والمنحدرة على الدخل X_2 والثروة X_3 بناء على 10 مفردات:[34]

$$\hat{Y}_i = 24.377 + 0.8716X_{2i} - 0.0349X_{3i} \qquad \textbf{(8.8.10)}$$
$$t = (3.875) \quad (2.7726) \quad (-1.1595) \qquad R^2 = 0.9682$$

معامل الثروة في هذا الانحدار ليس فقط لديه الإشارة الخاطئة، ولكنه أيضًا غير معنوي إحصائيًا عند مستوى معنوية 5%. ولكن عندما نزيد حجم العينة إلى 40 مشاهدة (التصغير الجزيئي؟)، سنحصل على النتائج التالية:

$$\hat{Y}_i = 2.0907 + 0.7299X_{2i} + 0.0605X_{3i} \qquad \textbf{(9.8.10)}$$
$$t = (0.8713) \quad (6.0014) \quad (2.0014) \qquad R^2 = 0.9672$$

الآن معامل الثروة ليس فقط لديه الإشارة الصحيحة والمتوقعة، ولكنه معنوي إحصائيًا عند مستوى معنوية 5%.

الحصول على بيانات إضافية أو بيانات "أفضل" قد لا يكون أمرًا سهلاً في أغلب الأحيان، وكما لاحظ Judge et al :

للأسف من الصعب حصول الباحث الاقتصادي على بيانات إضافية بدون تكلفة مرتفعة لاختيار القيم المرغوب فيها للمتغيرات المفسرة بالإضافة إلى ذلك، عندما يتم إضافة متغير جديد، فإنه يكون غير متحكم فيه، فيجب أن تتعامل بحذر مع البيانات المضافة التي كونها وبين البيانات الأخرى الموجودة في العينة الأصلية،

(34) خالص شكري لـ Albert Zucker لإعطائي النتائج الخاصة بهذه الانحدارات.

أي يجب أن تتأكد أن التكوين الاقتصادي الخاص بالمشاهدات الجديدة هو نفسه الخاص بالبيانات الأصلية[35].

6 - تقليل العلاقات الخطية في الانحدارات متعددة الحدود :
Reducing collinearity in polynomial regressions

في الفقرة 7.10 ناقشنا نماذج الانحدار متعددة الحدود. ومن خواص هذه النماذج أن المتغير أو المتغيرات المفسرة تظهر في صورة أسية. وبالتالي ففي دالة التكلفة التكعيبية والتي تشتمل على انحدار للتكلفة الكلية على الناتج، (الناتج) و(الناتج) كما هو (4.10.7)، فإن حدود الناتج المختلفة ستكون مرتبطة مما يجعل هناك صعوبة في تقدير معاملات الميل الخاص بها بشكل دقيق[36]. وبشكل عملي، فقد وجد أنه إذا تم التعبير عن المتغير أو المتغيرات المفسرة في صورة الانحرافات (أي الانحراف عن الوسط الحسابي) فإن تعدد العلاقات الخطية يقل بشكل ملحوظ. ولكن إذا ظلت المشكلة كما هي[37] فعلى الفرد أن يلجأ إلى أسلوب مثل متعددة الحدود المتعامدة[38].

7 - طرق أخرى لعلاج تعدد العلاقات الخطية :
Other methods of remedying multicollinearity

الأساليب الإحصائية متعددة المتغيرات مثل التحليل العاملي، وتحليل المكونات الرئيسية، أو أساليب مثل انحدار الحافة يتم استخدامها كثيراً " لحل " مشكلة تعدد العلاقات الخطية. للأسف هذه الأساليب خارج نطاق الدراسة لهذا الكتاب، حيث إنه لا يمكن مناقشتها بشكل جيد بدون الطرق إلى جبر المصفوفات[39].

---

(35) Judge et al., op. cit., p. 625. See also Sec. 10.9.

(36) كما لاحظنا، بما أن العلاقة X، X^2، X^3 غير خطية فإن الانحدار متعدد الحدود لا يخالف الفروض الخاصة بتعدد العلاقات الخطية في النموذج التقليدي.

(37) انظر R. A. Bradley and S. S. Srivastava, "Correlation and Polynomial Regression," American Statistician vol. 33, I 979, pp. I 1-14.

(38) انظر Norman Draper and Harry Smith, Applied Regression Analysis, 2d ed., Johm Wiley & Sons, New York, 1981, pp. 266 274.

(39) كتحليل منطقي ومهم لهذه الطرق من الناحية العملية ارجع إلى Chatterjee and Bertram Price, Regression Analysis by Example, John Wiley & Sons, New York, I 977, Chaps. 7 and 8. See also H. D. Vinod, "A Survey of Ridge Regression and Related Techniques for Improvements over Ordinary Least Squares," Review of Economics and Statistics vol.60, February 1978, pp. 121–131.

9.10 هل بالضرورة تعدد العلاقات الخطية يُعتبر أمراً سيئًا؟ ذلك ليس ضروريًا إذا كان الهدف هو التنبؤ فقط: IS MULTICOLLINEARITY NECESSARILY BAD? MAYBE NOT IF THE OBJECTIVE IS PREDICTION ONLY

وسبق وقيل إنه إذا كان الهدف الوحيد من تحليل الانحدار هو التنبؤ بالقيم المستقبلية، فإن تعدد العلاقات الخطية ليس مشكلة مهمة، حيث إنه كلما زادت R^2 تحسن التنبؤ[40].

ولكن ليصبح ذلك صحيحًا يجب أن " . . . طالما أن قيم المتغيرات المفسرة والمرغوب استخدامها في التنبؤ لها نفس قرب التبعية الخطية مثل التصميم الأصلي (للبيانات) للمصفوفة X "[41]

وبالتالي إذا وجد في تقديرات الانحدار أن $X_2 = 2X_3$ تقريبًا، بالتالي في عينة مستقبلية مستخدمة للتنبؤ بـ Y، X_2 يجب أن تكون أيضًا تقريبًا متساوية لـ $2X_3$ وهذا الشرط من الصعب تحقيقه عمليًا (انظر الهامش 35)، في مثل هذه الحالة يكون التنبؤ غير مؤكد بدرجة أكبر[42]. والأهم من ذلك، أنه إذا كان الهدف من التحليل لا يقتصر على التنبؤ فقط ولكن هناك حاجة للوصول إلى تقديرات مقبولة المعالم تصبح مشكلة تعدد العلاقات الخطية أكثر خطورة، حيث إننا رأينا أنها تؤدي إلى أخطاء قياسية كبيرة للمقدرات.

في بعض الأحيان، قد لا يؤدي تعدد العلاقات الخطية إلى مشكلة كبيرة.

هذه الحالة تحدث عندما يكون R^2 كبيرًا ومعاملات الانحدار معنوية انفراديًا مما يعني وجود قيم مرتفعة للـ t وعلى الرغم من ذلك قد نجد أن مؤشر الشرط يشير إلى

---

(40) انظر: R. C. Geary, "Some Results about Relations between Stochastic Variables: A Discusion Document," Review of Intemational Statistical Institute, vol. 31, 1963, pp. 163-181.

(41) Judge et al., op. cit., p. 619 . وستجد في هذه الصفحة أيضًا إثباتًا يعلل أنه على الرغم من وجود تعدد في العلاقات الخطية فإنه يمكن الحصول على تنبؤات الوسط أفضل إذا ظل التعدد في العلاقات الخطية على نفس الحال في العينات المستقبلية.

(42) لمناقشة جيدة في هذا الموضوع، انظر E. Malinvaud, Statistical Methods of Econometrics, 2d ed., North-Holland Publishing Company, Amsterdam, 1970, pp. 220–221.

وجود ارتباط متعدد قوى في البيانات. متى يمكن أن يحدث ذلك؟ الإجابة كما تصورها Johnston كالتالي:

يمكن أن يحدث ذلك إذا كانت معاملات الانحدار كلاً على حدة كبيرة عدديًا بصورة واضحة كالقيم الحقيقية، وبالتالي فإن التأثير سيظهر حتى وإن كانت الأخطاء القياسية متضخمة، وذلك بسبب أن القيمة الحقيقية نفسها كبيرة جداً حتى ولو كان المقدر غير سليم مازال له معنوية في التقدير[43].

10.10 مثال مطول.. بيانات Longley :
AN EXTENDED EXAMPLE: THE LONGLEY DATA

خلاصة هذا الفصل سيتم تقديمها بتحليل بيانات قام بتجميعها longley.[44] وعلى الرغم من أنه تم تجميعها في الأساس للتحقق من الدقة الحسابية لمقدرات المربعات الصغرى وفقًا لعدد من البرامج الإحصائية على الحاسب الآلي، إلا أن بيانات longley أصبحت معملاً لشرح العديد من مشاكل الاقتصاد القياسي بما فيها تعدد العلاقات الخطية. البيانات معطاة في جدول (7.10). البيانات تشمل سلاسل زمنية من العام 1947 حتى 1962 وتتضمن Y = عدد العمالة بالآلاف، $X_1 =$ GNP مقوم بالسعر، $X_2 =$ GNP بملايين الدولارات، X_3 = عدد الأفراد غير العاملين بالآلاف، X_4 = عدد الأفراد المسجلين على قوة الجيش، X_5 = تعدد غير رسمي للأفراد فوق 14 عامًا و X_6 = السنة وتساوي 1 في 1947 و 2 في 1948 و 16 في 1962.

جدول (7.10) بيانات Lengley

Observation	y	X_1	X_2	X_3	X_4	X_5	Time
1947	60,323	830	234,289	2356	1590	107,608	1
1948	61,122	885	259,426	2325	1456	108,632	2
1949	60,171	882	258,054	3682	1616	109,773	3
1950	61,187	895	284,599	3351	1650	110,929	4
1951	63,221	962	328,975	2099	3099	112,075	5
1952	63,639	981	346,999	1932	3594	113,270	6
1953	64,989	990	365,385	1870	3547	115,094	7
1954	63,761	1000	363,112	3578	3350	116,219	8
1955	66,019	1012	397,469	2904	3048	117,388	9
1956	67,857	1046	419,180	2822	2857	118,734	10
1957	68,169	1084	442,769	2936	2798	120,445	11

(43) J. Johnston, Econometric Methods, 3d ed., McGraw-Hill, New York, 1984, p. 249.
(44) Longley, J. "An Appraisal of Least-Squares Programs from the Point of the User," Journal of the American Statistical Association, vol. 62, 1967, pp. 819–481.

تابع – جدول (7.10) بيانات Lengley

Observation	y	X_1	X_2	X_3	X_4	X_5	Time
1959	68,655	1126	482,704	3813	2552	123,366	13
1960	69,564	1142	502,601	3931	2514	125,368	14
1961	69,331	1157	518,173	4806	2572	127,852	15
1962	70,551	1169	554,894	4007	2827	130,081	16

المصدر: انظر هامش رقم 44

افترض أن الهدف من التحليل هو التنبؤ بـ Y بناء على الستة متغيرات المفسرة X. باستخدام Eviews 3، حصلنا على نتائج الانحدار التالية:

Dependent Variable: *Y*
Sample: 1947-1962

Variable	Coefficient	Std. Error	*t*-Statistic	Prob.
C	-3482259.	890420.4	-3.910803	0.0036
X_1	15.06187	84.91493	0.177376	0.8631
X_2	-0.035819	0.033491	-1.069516	0.3127
X_3	-2.020230	0.488400	-4.136427	0.0025
X_4	-1.033227	0.214274	-4.821985	0.0009
X_5	-0.051104	0.226073	-0.226051	0.8262
X_6	1829.151	455.4785	4.015890	0.0030

R-squared	0.995479	Mean dependent var	65317.00
Adjusted R-squared	0.992465	S.D. dependent var	3511.968
S.E. of regression	304.8541	Akaike info criterion	14.57718
Sum squared resid	836424.1	Schwarz criterion	14.91519
Log likelihood	-109.6174	*F*-statistic	330.2853
Durbin-Watson stat	2.559488	Prob(*F*-statistic)	0.000000

في ضوء هذه النتائج، نستنتج أن هناك مشكلة علاقات خطية، حيث إن قيمة R^2 مرتفعة جداً. وعدد قليل من المتغيرات غير معنوي (X_1، X_2، X_5) وهذا يعتبر دليلاً معروفًا على وجود مشكلة تعدد العلاقات الخطية.

لتوضيح ذلك أكثر، دعنا نستعرض جدول (8.10) الذي يوضح الارتباطات التبادلية بين المتغيرات المنحدرة الستة.

جدول (8.10) الارتباطات التبادلية Inter correlations

	X_1	X_2	X_3	X_4	X_5	X_6
X_1	1.000000	0.991589	0.620633	0.464744	0.979163	0.991149
X_2	0.991589	1.000000	0.604261	0.446437	0.991090	0.995273
X_3	0.620633	0.604261	1.000000	−0.177421	0.686552	0.668257
X_4	0.464744	0.446437	−0.177421	1.000000	0.364416	0.417245
X_5	0.979163	0.991090	0.686552	0.364416	1.000000	0.993953
X_6	0.991149	0.995273	0.668257	0.417245	0.993953	1.000000

هذا الجدول يعطينا ما يسمى **بمصفوفة الارتباط**. فمدخلات هذا الجدول على القطر الرئيسي (الأرقام المعطاة من الطرف العلوي في اليسار إلى الطرف السفلي في اليمين) تمثل ارتباط المتغير مع نفسه، والتي تساوي دائمًا 1 وفقًا لتعريف الارتباط، والمدخلات الموجودة على جوانب القطر الرئيسي هي الارتباطات الثنائية بين المتغيرات X. إذا نظرنا إلى الصف الأول في هذا الجدول، فإنه يمثل ارتباط المتغير X_1 مع باقي المتغيرات. مثلاً 0.991589 تمثل الارتباط بين X_1 و X_2، 0.620633 هي الارتباط بين X_1 و X_3 وهكذا.

كما نرى، العديد من هذه الارتباطات الثنائية مرتفعة، مما يعني وجود مشكلة تعدد العلاقات الخطية. بالطبع تذكر التحذير الذي سبق وذكرناه والخاص بأن مثل هذه الارتباطات الثنائية قد تكون شرطًا كافيًا ولكن غير ضروري لوجود تعدد العلاقات الخطية.

لإلقاء مزيد من الضوء على طبيعة مشكلة تعدد العلاقات الخطية، دعنا نقوم بإجراء انحدارات مساعدة، أي انحدار لكل متغير X على باقي المتغيرات المفسرة. للاختصار سنستعرض فقط قيم R^2 التي حصلنا عليها من هذه الانحدارات، المعطاة في جدول (9.10). بما أن قيم R^2 في الانحدارات المساعدة كبيرة جدًا (مع استثناء انحدار X_4) على باقي المتغيرات X، يبدو أن لدينا بالفعل مشكلة تعدد في العلاقات الخطية. نفس المعلومة نحصل عليها باستخدام معاملات التفاوت. كما سبق وذكرنا، كلما اقترب معامل التفاوت من الصفر كلما كان ذلك دليلاً أقوى على وجود تعدد في العلاقات الخطية.

جدول (9.10) طبيعة مشكلة تعدد العلاقات الخطية

Dependent variable	R^2 value	Tolerance (TOL) = 1 − R^2
X_1	0.9926	0.0074
X_2	0.9994	0.0006
X_3	0.9702	0.0298
X_4	0.7213	0.2787
X_5	0.9970	0.0030
X_6	0.9986	0.0014

بتطبيق قاعدة Klein، سنرى أن قيم R^2 والتي حصلنا عليها من الانحدارات المساعدة تزيد عن قيمة R^2 الكلية (التي حصلنا عليها من انحدار Y على كل المتغيرات المفسرة X) بـ 0.9954 في 3 من كل 6 من الانحدارات المساعدة، مرة أخرى، تلك النتيجة تجعلنا نستنتج أنه بالتأكيد توجد مشكلة تعدد العلاقات الخطية في بيانات

longley . وعند تطبيق اختيار F المعطي في (3.7.10) سيجد القارئ أن كل قيم R^2 المعطاة في الجداول السابقة تختلف إحصائيًا عن الصفر .

لاحظنا سابقًا أن مقدرات OLS وأخطاءها القياسية حساسة جدًا لأي تغيير بسيط في البيانات . في تمرين (32.10) يطلب من القارئ أن يعيد انحدار Y على المتغيرات المفسرة X الستة ولكن مع إسقاط المشاهدة الأخيرة، أي عمل الانحدار في الفترة 1947 إلى 1961 . سنرى أن نتائج الانحدار تغيرت كثيرًا بمجرد إسقاط مشاهدة سنة واحدة فقط .

والآن بعد أن تأكدنا من وجود مشكلة تعدد في العلاقات الخطية، ما هو «العلاج» الذي يجب أن نقوم به؟ دعنا نعود مرة أخرى إلى النموذج الأصلي .

في البداية، يمكننا أن نعبر عن الـ GNP بحدود حقيقية وليست نوعية وذلك عن طريق قسمة GNP النوعي على معامل تصحيح الأسعار ثانيًا، بما أن عدد السكان غير الرسمي للأشخاص فوق 14 سنة يزيد مع مرور الزمن كطبق معدلات النمو التقليدية سيكون مرتبطًا بدرجة كبيرة مع متغير الزمن والذي يعبر عنه المتغير X_6 في النموذج السابق .

وبالتالي بدلاً من الاحتفاظ بالمتغيرين الاثنين معًا، سنحتفظ بالمتغير X_5 ونسقط X_6 . ثالثًا، لا يوجد سبب منطقي لاستخدام المتغير X_3 في النموذج وهو المتغير الذي يمثل عدد الأفراد غير العاملين ربما يكون معدل البطالة مقياسًا أفضل لشروط سوق العمالة . ولكن ليس لدينا بيانات عن المتغير الأخير . وبالتالي سنسقط المتغير X_3 . بعد القيام بكل هذه التغييرات، حصلنا على نتائج الانحدار التالية (GNP = RGNP الحقيقي)[45] :

Dependent Variable: Y
Sample: 1947-1962

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	65720.37	10624.81	6.185558	0.0000
RGNP	9.736496	1.791552	5.434671	0.0002
X_4	-0.687966	0.322238	-2.134965	0.0541
X_5	-0.299537	0.141761	-2.112965	0.0562

R-squared	0.981404	Mean dependent var	65317.00
Adjusted R-squared	0.976755	S.D. dependent var	3511.968
S.E. of regression	535.4492	Akaike info criterion	15.61641
Sum squared resid	3440470.	Schwarz criterion	15.80955
Log likelihood	-120.9313	F-statistic	211.0972
Durbin-Watson stat	1.654069	Prob(F-statistic)	0.000000

(45) معامل الارتباط بين X_5 و X_6 حوالي 0.9939 وهذا بالقطع يمثل ارتباطًا قويًا جدًا .

على الرغم من أن قيمة R^2 أقل قليلاً من قيمة R^2 الأصلية، إلا أنها مازالت قيمة كبيرة جداً. الآن كل المعاملات المقدرة معنوية وإشارة هذه المعاملات لها منطقية اقتصادية.

ستترك للقارئ إمكانية عمل تغييرات أخرى، ودراسة مدى تأثر وتغيير النتائج في كل مرة. مع الوضع في الاعتبار التحذير الذي سبق وأشارنا إليه والخاص باستخدام التحويل بالنسبة للتغلب على مشكلة تعدد العلاقات الخطية.

سنعود مرة أخرى إلى هذه النقط في (الفصل 11).

الخلاصة والنتائج : SUMMARY AND CONCLUSIONS

1 - أحد فروض نموذج الانحدار الخطي التقليدي عدم وجود تعدد في العلاقات الخطية بين المتغيرات المفسرة، الـ X's. فبشكل عام تعدد العلاقات الخطية يشير إلى الحالة التي تكون فيها علاقة خطية تامة أو غير تامة بين المتغيرات X.

2 - عواقب تعدد العلاقات الخطية هي كالتالي: إذا كان هناك تعدد كامل في العلاقات الخطية أو تام، فإن معاملات الانحدار تكون غير قابلة للتقدير وأخطائها القياسية غير معروفة. أما إذا كان هناك تعدد كبير في العلاقات الخطية ولكنه غير كامل، فإن تقدير معاملات الانحدار يكون ممكنًا ولكن أخطاءهم القياسية تكون كبيرة جداً. كنتيجة لذلك، فإن قيم معالم المجتمع لا يمكن تقديرها بدقة. عمومًا، إذا كان الهدف هو تقدير التوليفة الخطية المكونة من هذه المتغيرات، الدالة القابلة للتقدير، فإن ذلك من الممكن القيام به في وجود تعدد كامل للعلاقات الخطية.

3 - على الرغم من أنه لا توجد طرق أكيدة لاكتشاف تعدد العلاقات الخطية، إلا أن هناك مؤشرات متعددة لذلك كالتالي:

(a) أوضحت المؤشرات وجود R^2 كبيرة جدًا ولكن لا يوجد أي من معاملات الانحدار معنوي إحصائيًا باستخدام اختيار t. وهذه الحالة بالطبع حالة متطرفة.

(b) إذا كان النموذج لا يحتوي إلا على متغيرين اثنين فقط، فهذا دليل جيد على وجود تعدد في العلاقات الخطية يكون متحققًا إذا كان معامل الارتباط البسيط أو الصفري كبيرًا مما يعني وجود تعدد في العلاقات الخطية.

(c) ولكن معاملات الارتباط الصفرية قد تؤدي إلى نتيجة خاطئة إذا تم استخدامها في نماذج يوجد بها أكثر من متغيرين اثنين من الـ X ، حيث إنه يحدث أحيانًا أن نجد معاملات ارتباط صغيرة جدًا، ورغم ذلك يكون هناك تعدد قوى في العلاقات الخطية . في مثل هذه الحالات، لابد أن يلجأ الباحث إلى معاملات الارتباط الجزيئية .

(d) إذا كانت R^2 كبيرة، ولكن معاملات الارتباط الجزيئية صغيرة، قد يكون محتملاً وجود تعدد في العلاقات الخطية . فهنا يكون متغيرًا أو أكثر مزيفًا ولكن إذا كانت R^2 كبيرة ومعاملات الارتباط الجزيئية كبيرة أيضًا قد لا توجد مشكلة تعدد في العلاقات الخطية . أيضًا كما أوضح C.robert ، Krishna kumar ، John O'Hagan و Brendan McCabe هناك بعض المشكلات الإحصائية المرتبطة باختيار الارتباطات الجزيئية الذي اقترحه Farrar و Glauber .

(e) وبالتالي، فإنه يمكن للباحث أن يقوم بعمل انحدار لكل X على باقي المتغيرات المنحدرة X's الموجودة في النموذج، ويوجد في كل مرة معامل التحديد المرتبط بكل انحدار R_i^2 . وعندما يوجد R_i^2 كبير، فإن ذلك دليل على أن X_i مرتبطة بدرجة كبيرة مع باقي الـ X's . وهنا من الممكن إسقاط ذلك المتغير X_i من النموذج مع وضع في الاعتبار ألا يكون ذلك مصاحبًا لتحيز في التوصيف .

4 - اكتشاف تعدد العلاقات الخطية هو نصف الزجاجة . أما النصف الآخر هو كيفية التغلب على هذه المشكلة . مرة ثانية لا توجد طرق أكيدة لذلك، ولكن هناك بعض القواعد التي يمكن اتباعها . بعض هذه القواعد هي : (1) استخدام معلومات إضافية أو مسبقة، (2) مزج البيانات المقطعية ببيانات السلاسل الزمنية، (3) حذف المتغير الأكثر ارتباطًا، (4) تحويل البيانات و (5) الحصول على بيانات جديدة أو إضافية . بالطبع أي من هذه الطرق سيصلح لحل المشكلة عمليًا يعتمد بشكل كلي على طبيعة البيانات وحدة مشكلة تعدد العلاقات الخطية .

5 - سبق وذكرنا الدور الذي يلعبه تعدد العلاقات الخطية في التنبؤ، وأشرنا إلى أنه في حالة تكرار تكوين العلاقات الخطية في العينات المستقبلية، فإنه من الممكن استخدام مقدرات الانحدار، والتي تم التوصل إليها مع وجود تعدد في العلاقات الخطية، وتستخدم هذه المقدرات في أغراض التنبؤ .

6 - بالرغم من أن تعدد العلاقات الخطية نال اهتمامًا جيدًا في الدراسات الأدبية، فهناك مشكلة على نفس الخطورة ومتعلقة بالأبحاث التطبيقية وهي الخاصة بالتصغير الجزيئي والمقصود بها صغر حجم العينة. وبناء على ما قاله Goldberger فإن: " عندما يعاني بحث نظري من مشكلة تعدد العلاقات الخطية، فإن القارئ يرتب في معرفة ما إذا كانت هذه المعاناة ستحدث أيضًا إذا وجدت مشكلة "تصغير جزيئي" بدلاً من "تعدد العلاقات الخطية" أم لا[46].

وبالتالي، فإن Goldberger يقصد أن القارئ يرغب دائمًا في معرفة ما هو الحجم الذي يمكن أن نطلق عليه حجمًا صغيرًا للعينة والذي تنتج عنه مشاكل صغر حجم العينة كما رغب من قبل في معرفة المقدار الذي يجب أن نصل إليه حتى يمكننا أن نقول إن R^2 التي حصلنا عليها من أي من الانحدارات المساعدة كبيرة بدرجة كافية للقول بأن مشكلة تعدد العلاقات الخطية هي بالفعل مشكلة خطيرة في البيانات محل الدراسة.

تمارين: EXERCISES

أسئلة: Questions

1.10 في الانحدار الخطي الذي يشتمل على k من المتغيرات، يكون لدينا k من المعادلات الطبيعية لتقدير k من المجاهيل. هذه المعادلات الطبيعية معطاة في الملحق C. افترض أن X_k تعتبر توليفة خطية كاملة في باقي المتغيرات X. كيف يمكنك في هذه الحالة إيضاح أنه لا يمكن تقدير معاملات الانحدار (k معامل)؟

2.10 باستخدام البيانات الافتراضية الموجودة في جدول (10.10).

افترض أنك تريد تقدير النموذج التالي باستخدام هذه البيانات:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

(a) هل يمكنك تقدير الثلاثة مجاهيل؟ علل إجابتك.

(b) إذا كان ذلك غير ممكن، ما هي الدالة الخطية المكونة من هذه المعالم، (أي الدالة القابلة للتقدير)، والتي يمكن تقديرها؟ وضح كل الخطوات الحسابية اللازمة لذلك.

(46) Goldberger,op. cit., p. 250.

جدول (10.10)

Y	X_2	X_3
−10	1	1
−8	2	3
−6	3	5
−4	4	7
−2	5	9
0	6	11
2	7	13
4	8	15
6	9	17
8	10	19
10	11	21

3.10 بالرجوع إلى مثال وفيات الأطفال والذي استعرضناه في (الفصل 8) في هذا المثال، لدينا انحدار لمعدل وفيات الأطفال على الـ GNP بالنسبة للفرد (PGNP) ونسبة المتعلمين بين السيدات (FLR). والآن دعنا نفترض أننا أضفنا متغيراً جديداً وهو معدل الخصوبة (TFR) وحصلنا على النتائج التالية:

Dependent Variable: CM

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	168.3067	32.89165	5.117003	0.0000
PGNP	−0.005511	0.001878	−2.934275	0.0047
FLR	−1.768029	0.248017	−7.128663	0.0000
TFR	12.86864	4.190533	3.070883	0.0032

R-squared	0.747372	Mean dependent var	141.5000
Adjusted R-squared	0.734740	S.D. dependent var	75.97807
S.E. of regression	39.13127	Akaike info criterion	10.23218
Sum squared resid	91875.38	Schwarz criterion	10.36711
Log likelihood	−323.4298	F-statistic	59.16767
Durbin-Watson stat	2.170318	Prob(F-statistic)	0.000000

(a) قارن بين نتائج الانحدار الموجودة في المعادلة (1.2.8) والنتائج الحالية؟ وما هي المتغيرات الملحوظة؟ وما أثر تلك المتغيرات؟

(b) هل هناك فائدة من إضافة متغير TFR للنموذج؟ ولماذا؟

(c) بما أن كل معاملات الـ t المنفردة لها معنوية إحصائية، هل يمكن القول بأنه لا يوجد لدينا مشكلة علاقات خطية في المثال الحالي؟

4.10 إذا كانت لدينا العلاقة التالية: $\lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i} = 0$ متحققة لكل قيم λ_1، λ_2 و λ_3، قدر $r_{1\,2.3}$، $r_{1\,3.2}$ و $r_{2\,3.1}$. وأوجد أيضاً $R^2_{1.2\,3}$، $R^2_{2.1\,3}$ و $R^2_{3.1\,2}$. ما هي درجة تعدد العلاقات الخطية في هذه الحالة؟ لاحظ أن: $R^2_{1.2\,3}$ هو معامل التحديد في الانحدار الخاص بـ Y على X_2 و X_3. قيم R^2 الأخرى تعرف بنفس الطريقة.

5.10 اعتبر النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + \beta_4 X_{t-2} + \beta_5 X_{t-3} + \beta_6 X_{t-4} + u_t$$

حيث Y = الاستهلاك، X = الدخل و t = الزمن. النموذج السابق يفترض أن نفقات الاستهلاك عند الزمن t هي دالة ليس فقط في الدخل عند الزمن t ولكن أيضًا دالة في الدخل في فترات زمنية سابقة. وبالتالي فإن نفقات الاستهلاك في الربع الأول من عام 2000 تعتبر دالة في دخل نفس الربع السنوي والأربعة أرباع السنوية للعام 1999. مثل هذه النماذج تسمى نماذج توزيعية على فترات زمنية سابقة. وسنناقشها في فصل لاحق.

(a) هل تعتقد أن مثل هذه النماذج سيكون بها مشكلة تعدد العلاقات الخطية؟ ولماذا؟

(b) إذا توقعت وجود تعدد في العلاقات الخطية، كيف يمكنك حل هذه المشكلة؟

6.10 اعتبر النموذج التوضيحي الموجود في الفقرة 6.10. كيف يمكنك إعادة توفيق الفروق بين الميل الحدي إلى الاستهلاك الذي حصلنا عليه من (1.6.10) و (4.6.10)؟

7.10 عادة، توجد مشكلة تعدد العلاقات الخطية في البيانات التي تحتوي على سلاسل زمنية اقتصادية مثل GNP، المعروض من المال، الأسعار، الدخل، البطالة، وإلى ما غير ذلك. لماذا؟

8.10 اعتبر النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

حيث وجد أن r_{23}، وهو معامل الارتباط بين X_2 و X_3، يساوي الصفر. وبالتالي تم اقتراح عمل النماذج التالية:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + u_{1i}$$

$$Y_i = \gamma_1 + \gamma_3 X_{3i} + u_{2i}$$

(a) هل $\hat{\alpha}_2 = \hat{\beta}_2$ و $\hat{\gamma}_3 = \hat{\beta}_3$ ولماذا؟

(b) هل $\hat{\beta}_1$ ستساوي $\hat{\alpha}_1$ أو $\hat{\gamma}_1$ أو توليفة خطية منهما؟

(c) هل $\text{var}(\hat{\beta}_2) = \text{var}(\hat{\alpha}_2)$ و $\text{var}(\hat{\beta}_3) = \text{var}(\hat{\gamma}_3)$؟

9.10 بالعودة إلى المثال التوضيحي الموجود في (الفصل 7) والموجود به دالة إنتاج Cobb-Douglas والخاصة بالقطاع الزراعي التيواني. نتائج الانحدار معطاة في (4.9.7) وتوضح أن كلاً من معامل العمالة ورأس المال لهما معنوية إحصائية كل على حدة.

(a) هل متغيري العمالة ورأس المال مرتبطان بدرجة كبيرة أم لا؟

(b) إذا كان هناك ارتباط قوى بينهما، هل يمكن أن تسقط أحدهما، مثلاً العمالة من النموذج ونعمل انحداراً للناتج على رأس المال فقط؟

(c) إذا فعلاً قمت بذلك، ما هو تحيز التوصيف الذي قد ينتج عن ذلك؟ ادرس طبيعة هذا التحيز.

10.10 بالرجوع إلى مثال (4.7). افترض أن لديك مصفوفة الارتباط التالية:

	X_i	X_i^2	X_i^3
X_i	1	0.9742	0.9284
X_i^2		1.0	0.9872
X_i^3			1.0

(a) "بما أن معاملات الارتباط الصفرية كبيرة جداً، لابد أن يكون هناك تعدد في العلاقات الخطية". علق على ذلك.

(b) هل ستسقط X_i^2 و X_i^3 من النموذج؟

(c) إذا أسقطها، ما الذي سيحدث لمعامل X_i؟

11.10 الانحدار التدريجي. في أثناء عملية اختيار "أفضل" المتغيرات المفسرة لنموذج الانحدار، يتبع معظم الباحثين طريقة الانحدار التدريجي. هذه الطريقة تعتمد على إما إدخال متغير واحد فقط في النموذج في كل مرة أو إدخال كل المتغيرات X الممكنة في انحدار متعدد، ثم رفض كل منها على مدى (انحدار تدريجي خلفي). قرار الإضافة أو الحذف لأي متغير يعتبر على مدى اسهام هذا المتغير في الـ ESS، ويتم ذلك باستخدام اختيار F. بعد أن علمت الآن ما الذي يجب فعله في حالة وجود تعدد في العلاقات الخطية. ما هو الأسلوب الذي تفضل أن تتبعه؟ ولماذا؟(*)

(*) تأكد من أن تفكيرك يتفق مع Arthur S. Goldberg و D. B. Jochems, "Note on Stepwise Least-Squares," Journal of the American Statistical association, vol. 56, March 1961, pp. 105–110.

12.10 علل مع ذكر السبب ما إذا كانت العبارات التالية (صح أم خطأ أم غير متأكد منها):

(a) بغض النظر عن تعدد العلاقات الخطية. مقدرات OLS هي مقدرات BLUE.

(b) في حالة وجود تعدد كبير في العلاقات الخطية يكون من غير الممكن تحديد المعنوية المنفردة لكل معامل أو أكثر من معاملات الانحدار الجزيئي.

(c) إذا نتج عن أي انحدار مساعد قيمة كبيرة لـ R_i^2 فإن ذلك دليل قاطع على تعدد كبير في العلاقات الخطية.

(d) في حالة وجود معاملات ارتباط ثنائية كبيرة، فإن ذلك ليس دليلاً على وجود تعدد كبير في العلاقات الخطية.

(e) تعدد العلاقات الخطية يكون مشكلة غير حقيقية وغير ضارة بالتحليل إذا كان الهدف منه التنبؤ فقط.

(f) كلما زادت قيمة VIF كلما زاد تباين مقدرات الـ OLS.

(g) معامل التفاوت (TOL) مقياس أفضل للتعدد في العلاقات الخطية عن VIF.

(h) لن تحصل على قيمة كبيرة لـ R^2 في انحدار متعدد إذا كانت كل معاملات الميل الجزيئية غير معنوية إحصائيًا بناء على اختيار t المعتاد.

(i) في انحدار Y على X_2 و X_3، افترض أن لديك تباينًا صغيراً في قيم X_3 بما يزيد من تباين $(\hat{\beta}_3)$. في الحالة المتطرفة، إذا كانت كل الـ X_3 متساوية تمامًا فإن $\mathrm{Var}(\hat{\beta}_3)$ غير محدود.

13.10 (a) اثبت أنه إذا كان $r_{1i} = 0$ لكل من $i = 2$، 3، ...، k فإن

$$R_{1.2\,3\,\ldots\,k} = 0$$

(b) ما هي أهمية إيجاد انحدار المتغير $X_1(=Y)$ على X_1، X_3، ...، X_k؟

14.10 افترض أن كل معاملات الارتباط الصفرية لـ $X_1(=Y)$، X_2، ...، X_k تساوي r.

(a) ما هي قيمة $R_{21.\,23\,\ldots k}$؟

(b) ما هي قيم معاملات الارتباط من الدرجة الأولى؟

15.10[*] باستخدام المصفوفات ، من الممكن إثبات أن (انظر ملحق C) :

$$\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$$

(a) ما الذي يحدث لـ $\hat{\beta}$ إذا كان هناك ارتباط متعدد تام بين كل الـ X's؟

(b) كيف يمكنك تحديد ما إذا كان تعدد العلاقات الخطية التام موجود أم لا؟

16.10 [*] باستخدام المصفوفات، يمكن أثبات أن :

$$\text{var-cov}\ (\hat{\beta}) = \sigma^2(\mathbf{X'X})^{-1}$$

ما الذي يحدث لمصفوفة التباين - التغاير :

(a) عندما يكون هناك تعدد كامل في العلاقات الخطية؟

(b) عندما يكون تعدد العلاقات الخطية كبيرًا ولكنه ليس كاملاً؟

17.10 [*] اعتبر مصفوفة الارتباط التالية :

$$\mathbf{R} = \begin{array}{c} \\ X_2 \\ X_3 \\ \\ X_k \end{array} \begin{array}{c} \begin{array}{cccc} X_2 & X_3 & \cdots & X_k \end{array} \\ \begin{bmatrix} 1 & r_{23} & \cdots & r_{2k} \\ r_{32} & 1 & \cdots & r_{3k} \\ \cdots & \cdots & \cdots & \\ r_{k2} & r_{k3} & \cdots & 1 \end{bmatrix} \end{array}$$

كيف يمكنك باستخدام مصفوفة الارتباط تحديد (a) إذا كان هناك تعدد تام في العلاقات الخطية، (b) إذا كان هناك تعدد في العلاقات الخطية كبير ولكن ليس تامًا، (c) إذا كانت X's غير مرتبطة.

ملاحظة : من الممكن ان تستخدم $|\mathbf{R}|$ للإجابة على هذه الأسئلة، حيث إن $|\mathbf{R}|$ تشير إلى محدد الـ $\mathbf{R}$.

18.10[*] متغيرات مفسرة متعامدة. افترض أنه في النموذج التالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i$$

X_2 إلى X_k غير مرتبطين. هذه المتغيرات تسمى متغيرات متعامدة. إذا كان لدينا هذه الحالة :

(a) ما هو شكل المعنونة $(\mathbf{X'X})$؟

(b) كيف يمكنك الحصول على $\hat{\beta} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ ؟

(d) افترض أننا قمنا بعمل الانحدار ، وبعد ذلك أردت أن تضيف متغيرًا آخر متعامدًا، مثلاً X_{k+1} إلى النموذج. هل من الضروري أن نعيد عمل كل الحسابات السابقة مرة أخرى للحصول على تقديرات $\hat{\beta}_k$ إلى $\hat{\beta}_1$ ؟ علل إجابتك.

(*) اختياري.

19.10 اعتبر النموذج التالي :

$$GNP_t = \beta_1 + \beta_2 M_t + \beta_3 M_{t-1} + \beta_4 (M_t - M_{t-1}) + u_t$$

حيث GNP_t = GNP عند الزمن t، M_t = المعروض من المال عند الزمن t M_{t-1} = المعروض من المال عند الزمن $(t-1)$ و$(M_t - M_{t-1})$ = التغير في المعروض من المال بين الزمن t والزمن $t-1$. هذا النموذج يفترض أن مستويات الـ GUP عند الفريق t هو دالة في المعروض من المال عند الزمن t أيضًا و $(t-1)$ أيضًا بالإضافة إلى التغير في هذا المعروض بين هاتين الفترتين.

(a) إذا فرض وجود بيانات خاصة لتقدير هذا النموذج، هل ستنجح في تقدير معاملات هذا النموذج؟ علل إجابتك.

(b) إذا كانت الإجابة بلا، ما هي المعاملات التي يمكن تقديرها؟

(c) افترض عدم وجود المقدار $\beta_3 M_{t-1}$ في النموذج. هل إجابتك للجزء (a) ستظل كما هي؟

(d) كرر إجابة (c)، مع افتراض أن المقدار $\beta_2 M_t$ غير موجود في النموذج.

20.10 اثبت أن (7.4.7) و (8.4.7) يمكن إعادة كتابتهما كالتالي :

$$\hat{\beta}_2 = \frac{\left(\sum y_i x_{2i}\right)\left(\sum x_{3i}^2\right) - \left(\sum y_i x_{3i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right)\left(1 - r_{23}^2\right)}$$

$$\hat{\beta}_3 = \frac{\left(\sum y_i x_{3i}\right)\left(\sum x_{2i}^2\right) - \left(\sum y_i x_{2i}\right)\left(\sum x_{2i} x_{3i}\right)}{\left(\sum x_{2i}^2\right)\left(\sum x_{3i}^2\right)\left(1 - r_{23}^2\right)}$$

حيث r_{23} عبارة عن معامل الارتباط بين X_2 و X_3.

21.10 باستخدام (12.4.7) و (15.4.7). اثبت أنه في حالة وجود ارتباط متعدد تام، فإن تباين $\hat{\beta}_2$ و$\hat{\beta}_3$ غير محدود.

22.10 اثبت أن الأخطاء القياسية لمعاملات الميل المقدرة من (6.5.10) و (7.5.10) هي 0.1599 و 0.1825 على الترتيب (انظر الفقرة 5.10).

23.10 في نموذج الانحدار الذي يحتوي على k متغير من الممكن كتابة تباين معامل الانحدار الجزيئي k $(k = 2, 3, \ldots K)$ والموجود في (6.5.7) على الشكل التالي[*]

(*) هذه الصيغة معطاة في R. Stone, "The Analysis of Market Demand," Journal of the Royal Statistical Society, vol. B7, 1945, p. 297.
تذكر أيضًا (7.5.6)، للمزيد من التفاصيل انظر في Peter Kennedy, A Guide to Econometrics, 2d ed., The MIT Press, Cambridge, Mass., 1985, p. 156.

$$\text{var}(\hat{\beta}_k) = \frac{1}{n-k}\frac{\sigma_y^2}{\sigma_k^2}\left(\frac{1-R^2}{1-R_k^2}\right)$$

حيث σ_y^2 = تباين Y، σ_k^2 = تباين المتغير المفسر k، $R_k^2 = R^2$ التي نحصل عليها من انحدار X_k على باقي المتغيرات المفسرة X و R^2 = معامل التحديد من الانحدار المتعدد، أي انحدار Y على كل المتغيرات المفسرة X.

(a) بافتراض ثبات جميع المتغيرات الأخرى، إذا زادت σ_k^2 ما أثر ذلك على $\text{Var}(\hat{\beta}_k)$؟ وما أثر ذلك على مشكلة تعدد العلاقات الخطية؟

(b) ما الذي يحدث في المعادلة السابقة إذا كان هناك ارتباط متعدد تام؟

(c) صح أم خطأ : " تباين $\hat{\beta}_k$ يقل كلما زاد R^2 وبالتالي فتأثير R_k^2 ذات القيمة الكبيرة يتم تعادله مع قيمة R^2 الكبيرة.

24.10 من البيانات السنوية للقطاع الصناعي في الولايات المتحدة الأمريكية خلال الفترة 1899 إلى 1922، قام Dougherty بالحصول على نتائج الانحدار التالية :(*)

$$\begin{aligned} \widehat{\log Y} &= 2.81 - 0.53 \log K + 0.91 \log L + 0.047t \\ \text{se} &= (1.38) \quad (0.34) \qquad (0.14) \qquad (0.021) \qquad (1) \\ & R^2 = 0.97 \qquad F = 189.8 \end{aligned}$$

حيث Y = مؤشر الإنتاج الحقيقي، K = مؤشر مدخل رأس المال الحقيقي، L = مؤشر مدخل العمالة الحقيقي، t = الزمن أو الاتجاه العام.

باستخدام نفس البيانات، نحصل أيضًا على الانحدار التالي :

$$\begin{aligned} \widehat{\log (Y/L)} &= -0.11 + 0.11 \log (K/L) + 0.006t \\ \text{se} &= (0.03) \quad (0.15) \qquad (0.006) \qquad (2) \\ & R^2 = 0.65 \qquad F = 19.5 \end{aligned}$$

(a) هل هناك تعدد في العلاقات الخطية في الانحدار (1)؟ كيف عرفت؟

(b) في انحدار (1)، ما هي العلامة المتوقعة لـ $\log K$؟ هل النتائج متوافقة مع هذا التوقع؟ علل إجابتك.

(c) كيف يمكنك تفسير الشكل الدالي للانحدار (1)؟ (ملاحظة : استخدم دالة إنتاج Cobb–Douglas).

(d) فسر الانحدار (1). ما دور متغير الاتجاه العام في هذا الانحدار؟

(*) Christopher Dougherty, Introduction to Econometrics, Oxford University Press, New York, 1992, pp. 159–160.

(e) ما هو المنطق وراء تقدير نموذج الانحدار (2)؟

(f) إذا كان هناك تعدد في العلاقات الخطية في انحدار (1). ما هي القيود التي تم فرضها من خلال الباحث؟ (ملاحظة مساعدة: المقياس). كيف عرفت أن مثل هذا القيد من الممكن فرضه؟ ما الاختيار المستخدم؟ وضح كل خطواتك الحسابية.

(h) هل قيم R^2 التي حصلنا عليها من الانحدارين السابقين يمكن مقارنتهما؟ علل إجابتك. وإذا كان من الممكن مقارنتهما، كيف يمكنك القيام بذلك؟

25.10 علق على العبارات التالية بدقة:

(a) "في الحقيقة، مشكلة تعدد العلاقات الخطية ليست خطأ نمذجة، ولكنها حالة خاصة بالبيانات محل الدراسة" (*).

(b) "إذا لم يكن من الممكن الحصول على بيانات إضافية، فيجب على الفرد أن يتقبل الحقيقة الخاصة بالبيانات وأنها لا تعطينا إلا معلومات محدودة ويجب تبسيط النموذج تبعًا لذلك. فمحاولة تقدير نماذج معقدة بدرجة عالية أحد أهم الأخطاء التي يقع فيها الباحثون في مجال الاقتصاد القياسي التطبيقي" (†).

(c) "من المتعارف عليه بين الباحثين وجود مشكلة تعدد في العلاقات الخطية إذا كانت الإشارة المفترضة لمعامل الانحدار غير موجودة في النتائج العملية، أو إذا كانت بعض المتغيرات المفترض مسبقًا أهميتها في التحليل لا توجد لها قيم معنوية للـ t أو أهميتها إذا كانت نتائج الانحدار تتغير بشكل كبير عندما يتم حذف أحد المتغيرات المفسرة من النموذج. للأسف ولا واحد من هذه الشروط ضروري وكاف للجزم بوجود مشكلة تعدد العلاقة الخطية والأكثر من ذلك ولا واحد منها يعطي اقتراحًا مفيدًا حول طبيعة البيانات المطلوبة لحل مثل هذه المشكلة إذا وجدت فعلاً عند التطبيق" (‡).

(*) Samprit Chatterjee, Ali S. Hadi, and Betram Price, Regression Analysis by Example, 3d ed., John Wiley & Sons, New York, 2000, p. 226.

(†) Russel Davidson and James G. MacKinnon, Estimation and Inference in Econometrics, Oxford University Press, New York, 1993, p. 186.

(‡) Peter Kennedy, A Guide to Econometrics, 4th ed., MIT Press, Cambridge, Mass., 1998, p. 187

(d) " . . . . أي انحدار لسلسلة زمنية يحتوى على أكثر من أربعة متغيرات مستقلة لا يمكن الاعتماد على نتائجه مطلقًا "(*).

مسـائـل : Problems

26.10 Klein و Goldberger يحاولان تقدير نماذج الانحدار التالي وفقًا للاقتصاد الأمريكي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + u_i$$

حيث Y = الاستهلاك، X_2 = دخل العاملين بأجر، X_3 = دخل غير العاملين بأجر وغير العاملين بالزراعة، X_4 = دخل العاملين بالزراعة . ولكن بما أن X_2، X_3 و X_4 من المتوقع أن يكون بينها تعدد كبير في العلاقات الخطية فإنه تم تقدير β_3 و β_4 من تحليل بيانات مقطعية كالتالي :

β_2 $\beta_3 = 0.75$ و β_1 $\beta_4 = 0.625$. وباستخدام هذه المقدرات تم إعادة صياغة دالة الاستهلاك كالتالي :

$$Y_i = \beta_1 + \beta_2(X_{2i} + 0.75\ X_{3i} + 0.625\ X_{4i}) + u_i = \beta_1 + \beta_2 Z_i + u_i$$

حيث

$$Z_i = X_{2i} + 0.75\ X_{3i} + 0.625\ X_{4i}$$

(a) وفق النموذج المعدل للبيانات الموجودة في جدول (11.10) واحصل على تقديرات للمعاملات من β_1 إلى β_4 .

(b) كيف يمكنك تفسير المتغير Z؟

27.10 جدول (12.10) يعطي البيانات الخاصة بالصادرات، GDP، مؤشر سعر المستهلك (CPI) في الفترة 1970 إلى 1998 . دعنا نعتبر النموذج التالي :

$$\ln \text{Imports}_t = \beta_1 + \beta_2 \ln \text{GDP}_t + \beta_3 \ln \text{CPI}_t + u_t$$

(a) قدر معاملات هذا النموذج باستخدام البيانات المعطاة في الجدول .

(b) هل تعتقد أن هناك مشكلة تعدد في العلاقات الخطية في هذه البيانات؟

(*) هذه العبارة تعود إلى عالم الاقتصاد القياسي القديم Zvi Griliches، وموجودة في كتاب Ernst R. Berndt، تطبيقات الاقتصاد القياسي : Classic and Contemporary, Addion Wesley, Reading, Mass., 1991, p. 224.

جدول (11.10)

Year	Y	X_2	X_3	X_4	Year	Y	X_2	X_3	X_4
1936	62.8	43.41	17.10	3.96	1946	95.7	76.73	28.26	9.76
1937	65.0	46.44	18.65	5.48	1947	98.3	75.91	27.91	9.31
1938	63.9	44.35	17.09	4.37	1948	100.3	77.62	32.30	9.85
1939	67.5	47.82	19.28	4.51	1949	103.2	78.01	31.39	7.21
1940	71.3	51.02	23.24	4.88	1950	108.9	83.57	35.61	7.39
1941	76.6	58.71	28.11	6.37	1951	108.5	90.59	37.58	7.98
1945*	86.3	87.69	30.29	8.96	1952	111.4	95.47	35.17	7.42

* البيانات الخاصة بسنوات الحروب 1942 إلى 1944 مفقودة. بيانات السنوات الأخرى مقدرة بالبليون دولار لسنة 1939.

المصدر: L. R. Klein and A. S. Goldberger, An Economic Model of the United States, 1929–1952, North Holland Publishing Company, Amsterdam, 1964, p. 131.

جدول (12.10) صادرات الولايات المتحدة الأمريكية، GDP و CPI للفترة 1970 – 1998

Observation	CPI	GDP	Imports	Observation	CPI	GDP	Imports
1970	38.8	1039.7	39,866	1985	107.6	4213.0	338,088
1971	40.5	1128.6	45,579	1986	109.6	4452.9	368,425
1972	41.8	1240.4	55,797	1987	113.6	4742.5	409,765
1973	44.4	1385.5	70,499	1988	118.3	5108.3	447,189
1974	49.3	1501.0	103,811	1989	124.0	5489.1	477,365
1975	53.8	1635.2	98,185	1990	130.7	5803.2	498,337
1976	56.9	1823.9	124,228	1991	136.2	5986.2	490,981
1977	60.6	2031.4	151,907	1992	140.3	6318.9	536,458
1978	65.2	2295.9	176,002	1993	144.5	6642.3	589,441
1979	72.6	2566.4	212,007	1994	148.2	7054.3	668,590
1980	82.4	2795.0	249,750	1995	152.4	7400.5	749,574
1981	90.9	3131.3	265,067	1996	156.9	7813.2	803,327
1982	96.5	3259.2	247,642	1997	160.5	8300.8	876,366
1983	99.6	3534.9	268,901	1998	163.0	8759.9	917,178
1984	103.9	3932.7	332,418				

(c) أوجد انحدار:

$$\ln \text{Imports}_t = A_1 + A_2 \ln \text{GDP}_t \quad (1)$$

$$\ln \text{Imports}_t = B_1 + B_2 \ln \text{GDI}_t \quad (2)$$

$$\ln \text{GDP}_t = C_1 + C_2 + \ln \text{CPI}_t \quad (3)$$

وبناء على هذه الانحدارات، ما الذي تستنتجه عن تعدد العلاقات الخطية في هذه البيانات؟

(d) بافتراض وجود تعدد في العلاقات الخطية في هذه البيانات، ولكن $\hat{\beta}_2$ و $\hat{\beta}_3$ لهما معنوية منفردة عند مستوى معنوية 5% واختيار F الكلي معنوي أيضًا. هل في مثل هذه الحالة يفترض أن تقلق بشأن وجود مشكلة تعدد العلاقات الخطية أم لا؟

28.10 بالعودة إلى تمرين 19.7 والخاص بدالة الطلب على الدواجن في الولايات المتحدة.

(a) باستخدام نموذج خطي لوغاريتمي أو لوغاريتمي مزدوج قدر الانحدارات المساعدة ما عددها؟

(b) من نتائج الانحدارات المساعدة، كيف يمكن أن تقدر أي من المتغيرات المفسرة المرتبطة بشكل تعددي كبير؟ ما الاختيار المستخدم؟ وضح كل خطواتك الحسابية.

(c) إذا كان هناك تعدد في العلاقات الخطية معنوي في هذه البيانات، أي المتغيرات يجب أن يحذف لتقليل حدة هذه المشكلة؟ وإذا قمت بذلك فعلاً، ما المشكلة الاقتصادية التي يمكن أن تواجهها؟

(b) هل لديك أي اقتراحات بخلاف حذف المتغيرات للتعامل مع مشكلة تعدد العلاقات الخطية؟ وضح ذلك.

29.10 جدول 13.10 يعطي البيانات الخاصة بمبيعات السيارات الجديدة في الولايات المتحدة كدالة في متغيرات متعددة أخرى.

(a) كون نموذجًا خطيًا أو خطيًا لوغاريتميًا مناسبًا لتقدير دالة الطلب على السيارات في الولايات المتحدة.

(b) إذا قررت إدخال كل المتغيرات المنحدرة الموجودة في الجدول كمتغيرات مفسرة في النموذج. هل تعتقد أنك ستواجه مشكلة تعدد العلاقات الخطية؟ ولماذا؟

(c) إذا قمت بذلك فعلاً، كيف ستحصل هذه المشكلة؟ استعرض فروضك بشكل محدد ووضح كل خطواتك الحسابية بدقة.

جدول (13.10)

Year	Y	X_2	X_3	X_4	X_5	X_6
1971	10,227	112.0	121.3	776.8	4.89	79,367
1972	10,872	111.0	125.3	839.6	4.55	82,153
1973	11,350	111.1	133.1	949.8	7.38	85,064
1974	8,775	117.5	147.7	1,038.4	8.61	86,794
1975	8,539	127.6	161.2	1,142.8	6.16	85,846
1976	9,994	135.7	170.5	1,252.6	5.22	88,752
1977	11,046	142.9	181.5	1,379.3	5.50	92,017
1978	11,164	153.8	195.3	1,551.2	7.78	96,048
1979	10,559	166.0	217.7	1,729.3	10.25	98,824
1980	8,979	179.3	247.0	1,918.0	11.28	99,303
1981	8,535	190.2	272.3	2,127.6	13.73	100,397
1982	7,980	197.6	286.6	2,261.4	11.20	99,526

تابع - جدول (13.10)

Year	Y	X_2	X_3	X_4	X_5	X_6
1982	7,980	197.6	286.6	2,261.4	11.20	99,526
1983	9,179	202.6	297.4	2,428.1	8.69	100,834
1984	10,394	208.5	307.6	2,670.6	9.65	105,005
1985	11,039	215.2	318.5	2,841.1	7.75	107,150
1986	11,450	224.4	323.4	3,022.1	6.31	109,597

Y = مبيعات السيارات الجديدة (بالآلاف) غير مخلصة من الأثر الموسمي.

X_2 = مؤشر سعر المستهلك للسيارات الجديدة، 1967=100، غير مخلصة من الأثر الموسمي.

X_3 = مؤشر سعر المستهلك، كل المتغيرات، لكل مستهلكين المدينة، 1967=100 غير مخلصة من الأثر الموسمي.

X_4 = الدخل الشخصي المتغير (PDI)، بليون دولار، غير مخلصة من الأثر الموسمي.

X_5 = معدل الفائدة، نسبة محددة مباشرة من الشركات المالية.

X_6 = قوة العمل المدنية (بالآلاف)، غير مخلصة من الأثر الموسمي.

المصدر: Business Statistics, 1986, A Supplement to the Current Survey of Business, U.S. Department of Commerce.

30.10 لدراسة إمكانية ضمان أجر سنوي (ضرائب الدخل السالب)، قامت مؤسسة Rand بعمل دراسة للتعرف على أثر زيادة الأجر بالساعة على المعروض من العمالة (متوسط ساعات العمل)(*). بيانات هذه الدراسة مأخوذة من عينة قومية من 6000 أسرة يكون فيها رب الأسرة دخله أقل من 15,000 دولار سنويًا. البيانات مقسمة إلى 39 منطقة جغرافية لتنشيط عملية التحليل. هذه البيانات موجودة في جدول (14.10). وحيث إن هناك بيانات خاصة بأربع مناطق جغرافية مفقودة لبعض المتغيرات، فالبيانات المعطاة في الجدول تستعرض فقط 35 منطقة جغرافية. تعريفات المتغيرات المختلفة المستخدمة في التحليل معطاة في نهاية الجدول.

(a) قم بعمل انحدار لساعات العمل خلال العام على المتغيرات المعطاة في الجدول وفسر انحدارك.

(b) هل هناك دليل على وجود مشكلة تعدد في العلاقات الخطية في هذه البيانات؟ كيف عرفت ذلك؟

(*)D. H. Greenberg and M. Kosters, Income Guarantees and the working Poor, Rand Corporation, R-579-OEO, December 1970.

(c) احسب معامل تضخم التباين (VIF) و TOL لكل المتغيرات المقدرة.

(d) إذا كانت هناك مشكلة تعدد في العلاقات الخطية، ما هي الخطوة العلاجية التي يمكنك القيام بها (إذا كان هناك علاج)؟

(e) ما الذي تستنتجه من هذه الدراسة بخصوص إمكانية وجود ضرائب الدخل السالب؟

31.10 جدول (15.10) يعطي بيانات عن معدل الجريمة في 47 ولاية في الولايات المتحدة الأمريكية عام 1960. حاول أن تكون نموذجًا مناسبًا لتفسير معدل الجريمة وعلاقته بالمتغيرات الاجتماعية الاقتصادية الـ14 المعطاة في الجدول. ضع في اعتبارك أثناء تكوين هذا النموذج مشكلة تعدد العلاقات الخطية.

جدول (14.10) عدد ساعات العمل وبيانات أخرى لـ 35 مجموعة

Hours of work and atber Data for 35 graups

Observation	Hours	Rate	ERSP	ERNO	NEIN	Assets	Age	DEP	School
1	2157	2.905	1121	291	380	7250	38.5	2.340	10.5
2	2174	2.970	1128	301	398	7744	39.3	2.335	10.5
3	2062	2.350	1214	326	185	3068	40.1	2.851	8.9
4	2111	2.511	1203	49	117	1632	22.4	1.159	11.5
5	2134	2.791	1013	594	730	12710	57.7	1.229	8.8
6	2185	3.040	1135	287	382	7706	38.6	2.602	10.7
7	2210	3.222	1100	295	474	9338	39.0	2.187	11.2
8	2105	2.493	1180	310	255	4730	39.9	2.616	9.3
9	2267	2.838	1298	252	431	8317	38.9	2.024	11.1
10	2205	2.356	885	264	373	6789	38.8	2.662	9.5
11	2121	2.922	1251	328	312	5907	39.8	2.287	10.3
12	2109	2.499	1207	347	271	5069	39.7	3.193	8.9
13	2108	2.796	1036	300	259	4614	38.2	2.040	9.2
14	2047	2.453	1213	297	139	1987	40.3	2.545	9.1
15	2174	3.582	1141	414	498	10239	40.0	2.064	11.7
16	2067	2.909	1805	290	239	4439	39.1	2.301	10.5
17	2159	2.511	1075	289	308	5621	39.3	2.486	9.5
18	2257	2.516	1093	176	392	7293	37.9	2.042	10.1
19	1985	1.423	553	381	146	1866	40.6	3.833	6.6
20	2184	3.636	1091	291	560	11240	39.1	2.328	11.6
21	2084	2.983	1327	331	296	5653	39.8	2.208	10.2
22	2051	2.573	1194	279	172	2806	40.0	2.362	9.1
23	2127	3.262	1226	314	408	8042	39.5	2.259	10.8
24	2102	3.234	1188	414	352	7557	39.8	2.019	10.7
25	2098	2.280	973	364	272	4400	40.6	2.661	8.4
26	2042	2.304	1085	328	140	1739	41.8	2.444	8.2
27	2181	2.912	1072	304	383	7340	39.0	2.337	10.2
28	2186	3.015	1122	30	352	7292	37.2	2.046	10.9
29	2188	3.010	990	366	374	7325	38.4	2.847	10.6
30	2077	1.901	350	209	95	1370	37.4	4.158	8.2
31	2196	3.009	947	294	342	6888	37.5	3.047	10.6
32	2093	1.899	342	311	120	1425	37.5	4.512	8.1

تابع – جدول (14.10) عدد ساعات العمل وبيانات أخرى لـ 35 مجموعة

Observation	Hours	Rate	ERSP	ERNO	NEIN	Assets	Age	DEP	School
33	2173	2.959	1116	296	387	7625	39.2	2.342	10.5
34	2179	2.971	1128	312	397	7779	39.4	2.341	10.5
35	2200	2.980	1126	204	393	7885	39.2	2.341	10.6

لاحظ أن:

Hours = متوسط ساعات العمل خلال العام.

Rate = متوسط الأجر بالساعة (بالدولار).

ERSP = متوسط المكسب السنوي للزوج / الزوجة (بالدولار).

ERNO = متوسط المكسب السنوي لأفراد الأسرة الآخرين (بالدولار).

NEIN = متوسط الدخل السنوي (أي دخل بخلاف الأجر).

Assets = متوسط أصول الأسرة (حسابات بنكية وخلافه) (بالدولار).

Age = متوسط عمر المبحوث.

Dep = متوسط عدد الأفراد الذين يعولهم المبحوث.

School = متوسط أعلى شهادة تعليمية تم الحصول عليها.

المصدر: D. H. Greenberg and M. Kosters, Income Guarantees and the Working Poor, The Rand Corporation, R-579-OEO, December 1970.

32.10 بالرجوع إلى بيانات langley المعطاة في الفقرة 10.10. قم بتكرار الانحدار المعطي في الجدول بعد حذف بيانات عام 1962 أي قم بعمل الانحدار على الفترة 1947 إلى 1961. قارن الانحدارين، وما الاستنتاج العام الذي توصلت إليه من هذا التمرين؟

جدول (15.10) بيانات الجريمة في الولايات المتحدة والخاصة بـ 47 ولاية في عام 1960

U.S. Crime Data for 47 states in 1960

Observation	R	Age	S	ED	EX_0	EX_1	LF	M	N	NW	U_1	U_2	W	X
1	79.1	151	1	91	58	56	510	950	33	301	108	41	394	261
2	163.5	143	0	113	103	95	583	1012	13	102	96	36	557	194
3	57.8	142	1	89	45	44	533	969	18	219	94	33	318	250
4	196.9	136	0	121	149	141	577	994	157	80	102	39	673	167
5	123.4	141	0	121	109	101	591	985	18	30	91	20	578	174
6	68.2	121	0	110	118	115	547	964	25	44	84	29	689	126
7	96.3	127	1	111	82	79	519	982	4	139	97	38	620	168
8	155.5	131	1	109	115	109	542	969	50	179	79	35	472	206
9	85.6	157	1	90	65	62	553	955	39	286	81	28	421	239
10	70.5	140	0	118	71	68	632	1029	7	15	100	24	526	174
11	167.4	124	0	105	121	116	580	966	101	106	77	35	657	170
12	84.9	134	0	108	75	71	595	972	47	59	83	31	580	172
13	51.1	128	0	113	67	60	624	972	28	10	77	25	507	206
14	66.4	135	0	117	62	61	595	986	22	46	77	27	529	190

تابع – جدول (15.10) بيانات الجريمة في الولايات المتحدة والخاصة بـ 47 ولاية في عام 1960

Observation	R	Age	S	ED	EX_0	EX_1	LF	M	N	NW	U_1	U_2	W	X
15	79.8	152	1	87	57	53	530	986	30	72	92	43	405	264
16	94.6	142	1	88	81	77	497	956	33	321	116	47	427	247
17	53.9	143	0	110	66	63	537	977	10	6	114	35	487	166
18	92.9	135	1	104	123	115	537	978	31	170	89	34	631	165
19	75.0	130	0	116	128	128	536	934	51	24	78	34	627	135
20	122.5	125	0	108	113	105	567	985	78	94	130	58	626	166
21	74.2	126	0	108	74	67	602	984	34	12	102	33	557	195
22	43.9	157	1	89	47	44	512	962	22	423	97	34	288	276
23	121.6	132	0	96	87	83	564	953	43	92	83	32	513	227
24	96.8	131	0	116	78	73	574	1038	7	36	142	42	540	176
25	52.3	130	0	116	63	57	641	984	14	26	70	21	486	196
26	199.3	131	0	121	160	143	631	1071	3	77	102	41	674	152
27	34.2	135	0	109	69	71	540	965	6	4	80	22	564	139
28	121.6	152	0	112	82	76	571	1018	10	79	103	28	537	215
29	104.3	119	0	107	166	157	521	938	168	89	92	36	637	154
30	69.6	166	1	89	58	54	521	973	46	254	72	26	396	237
31	37.3	140	0	93	55	54	535	1045	6	20	135	40	453	200
32	75.4	125	0	109	90	81	586	964	97	82	105	43	617	163
33	107.2	147	1	104	63	64	560	972	23	95	76	24	462	233
34	92.3	126	0	118	97	97	542	990	18	21	102	35	589	166
35	65.3	123	0	102	97	87	526	948	113	76	124	50	572	158
36	127.2	150	0	100	109	98	531	964	9	24	87	38	559	153
37	83.1	177	1	87	58	56	638	974	24	349	76	28	382	254
38	56.6	133	0	104	51	47	599	1024	7	40	99	27	425	225
39	82.6	149	1	88	61	54	515	953	36	165	86	35	395	251
40	115.1	145	1	104	82	74	560	981	96	126	88	31	488	228
41	88.0	148	0	122	72	66	601	998	9	19	84	20	590	144
42	54.2	141	0	109	56	54	523	968	4	2	107	37	489	170
43	82.3	162	1	99	75	70	522	996	40	208	73	27	496	224
44	103.0	136	0	121	95	96	574	1012	29	36	111	37	622	162
45	45.5	139	1	88	46	41	480	968	19	49	135	53	457	249
46	50.8	126	0	104	106	97	599	989	40	24	78	25	593	171
47	84.9	130	0	121	90	91	623	1049	3	22	113	40	588	160

تعريفات المتغيرات :

R = معدل الجريمة، عدد الحوادث التي تم التبليغ عنها في الشرطة بالنسبة لكل مليون من السكان.

Age = عدد الذكور في العمر 14 إلى 24 لكل 1000 من السكان.

S = متغير وهمي للولايات الجنوبية (0 = لا، 1 = نعم).

ED = متوسط عدد سنوات التعليم مضروبة في 10 للأشخاص الذين تكون أعمارهم 25 فأكثر.

EX_0 = النفقات الموجهة للشرطة من خلال الولاية والحكومة المحلية للفرد في 1960.

EX_1 = النفقات الموجهة للشرطة من خلال الولاية والحكومة المحلية للفرد في 1959.

LF = معدل مشاركة قوة العمل لكل 1000 مواطن ذكر في المدينة.

M = عدد الذكور لكل 1000 أنثى.

N = حجم السكان في الولاية بمئات الآلاف.

NW = عدد السكان (ليسوا ذوي البشرة البيضاء) لكل 1000 من السكان.

U_1 = معدل البطالة بين الذكور في المدينة لكل 1000 في العمر 14 إلى 24.

U_2 = معدل البطالة بين الذكور في المدينة لكل 1000 في العمر 35 إلى 39.

W = وسيط أصول أو دخل الأسرة المحول بعشرات الدولارات.

X = عدد الأسر لكل 1000 يكون مكسبهم نصف وسيط الدخل.

ملاحظة: الولايات (47 ولاية في العام 1960).

المصدر: W. Vandaele, "Participation in Illegitimate Activities: Erlich Revisted, "in A. Blumstein, J. Cohen, and Nagin, D., eds., Deterrence and Incapapcitation, National Academy of Sciences, 1978, pp. 270–335.

الفصل الحادي عشر

اختلاف التباين: ماذا يحدث إذا كان تباين الخطأ غير ثابت؟

HETEROSCEDASTICITY: WHAT HAPPENS IF THE ERROR VARIANCE IS NONCONSTANT?

من الفروض المهمة لنموذج الانحدار الخطي التقليدي (فرض 4) أن يكون مقدار الخطأ u_i، والموجود في دالة انحدار المجتمع ثابت التباين، أي أن يكون لها جميعًا نفس قيمة التباين. في هذا الفصل، سنختبر صحة هذا الفرض، ونستعرض ماذا يحدث عندما لا يتحقق هذا الفرض؟. وكما في (الفصل 10)، نحن نحتاج إجابات على الأسئلة التالية:

1 – ما هي طبيعة مشكلة اختلاف التباين؟

2 – ما هي نتائجها؟

3 – كيف يمكن التعرف عليها؟

4 – ما هي الإجراءات العلاجية الممكنة؟

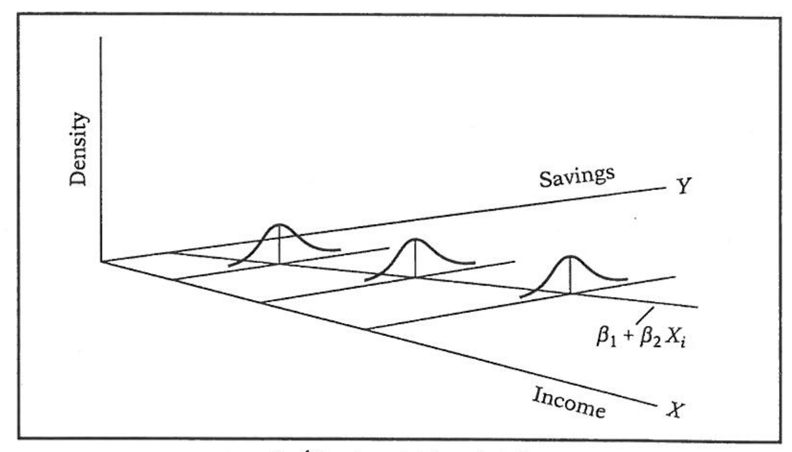
1.11 طبيعة اختلاف التباين:

THE NATURE OF HETEROSCEDASTICITY

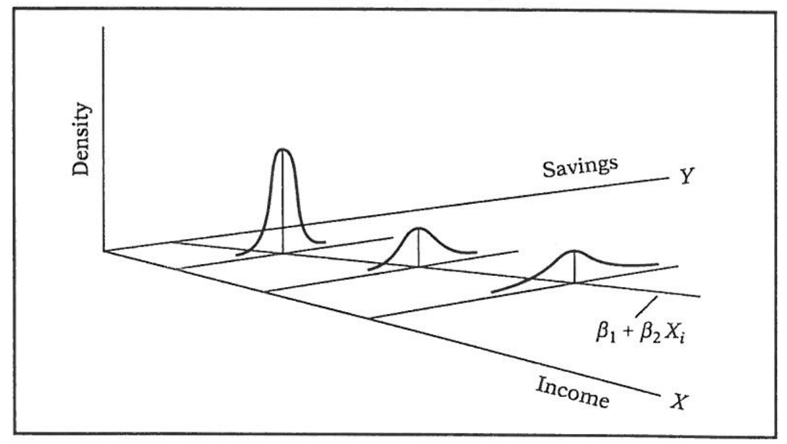
كما لاحظنا في (الفصل 3)، واحدًا من أهم فروض نموذج الانحدار الخطي التقليدي، هو أن يكون تباين مقدار الخطأ u_i، مشروطًا بالقيمة المحددة للمتغيرات المفسرة، مقدار ثابت يساوي σ^2. هذا هو فرض ثبات التباين (homoscedasticity)، أو تساوي (homo) الانتشار (scedasticity) أي تساوي التباين. بالرموز يكون لدينا:

$$E(u_i^2) = \sigma^2 \qquad i = 1, 2, ..., n \qquad (1.1.11)$$

بيانيًا، في نموذج انحدار ثنائي المتغيرات، يمكن توضيح ثبات التباين في الشكل (4.3) ولسهولة المتابعة تم إعادة عرض نفس الشكل البياني في الشكل (1.11).



شكل (1.11) ثبات تباين الأخطاء



شكل (2.11) اختلاف تباين الأخطاء

كما يوضح الشكل (1.11)، التباين المشروط لـ Y_i (والذي يساوي u_i)، مشروط بالقيمة المعطاة لـ X_i، تظل نفس القيمة بغض النظر عن القيمة التي يأخذها المتغير X.

علي النقيض في الشكل (2.11)، والذي يوضح التباين المشروط لـ Y_i أنه يتزايد كلما زادت X. هنا تباين Y_i غير ثابت.

وبالتالي، هناك اختلاف في التباين، بالرموز يكون كالتالي:

$$E(u_i^2) = \sigma_i^2 \qquad (2.1.11)$$

لاحظ أن الترقيم السفلي لـ σ^2، والخاص بالتباين المشروط لـ u_i (تباين مشروط لـ Y_i) لم يعد مقدارًا ثابتًا.

لمعرفة الفرق بين ثبات التباين واختلاف التباين، اقتضى أن النموذج ثنائي المتغيرات هو: $Y_i = \beta_1 + \beta_2 X_i + u_i$، Y_i، Y تمثل الادخار، و X تمثل الدخل. الشكلان (1.11) و(2.11) يوضحان أنه كلما زاد الدخل، فإن الادخار في المتوسط يزداد أيضًا. ولكن في شكل (1.11) تباين الادخار يظل كما هو لكل مستويات الدخل، ولكن في شكل (2.11) يزداد التباين مع الدخل. فمن الشكل (2.11) يبدو أنه كلما زاد في المتوسط دخل الأسر ذات الدخل المرتفع كلما زاد ادخارهم عن الأسر ذات الدخل المنخفض، ولكن هناك أيضًا تنوعًا كبيرًا في الادخار.

هناك أسباب عديدة تجعل تباين u_i متغيرًا. بعض من هذه الأسباب كالتالي:[1]

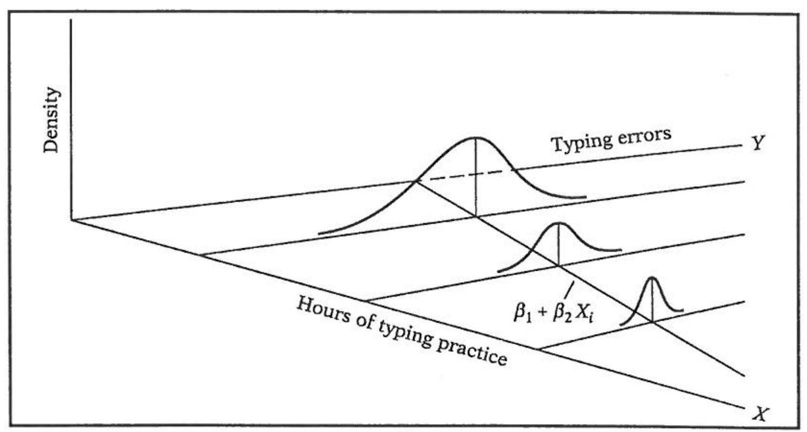
1 - وفقًا لنموذج التعلم من الأخطاء، فإن الإنسان يتعلم، والأخطاء تقل بمرور الزمن. في هذه الحالة، فإن من المتوقع أن تقل σ_i^2. وكمثال لذلك، اعتبر الشكل (3.11)، والذي يربط بين عدد الأخطاء المطبعية والحادثة في زمن محدد في اختبار ما وعدد ساعات التدريب على الطباعة. وكما يتضح من الشكل (3.11) كلما زاد عدد ساعات التدريب على الطباعة كلما قل في المتوسط عدد الأخطاء المطبعية، وكلما قل أيضًا تباينها.

2 - كلما يزداد الدخل، كلما يزداد تنوع صرف الدخل[2]، وتكون هناك مجالات أوسع لصرف الأموال. وبالتالي، فإنه من المتوقع أن تزداد σ_i^2 مع زيادة الدخل، وبالتالي في انحدار الادخار على الدخل ممكن أن نرى زيادة σ_i^2 مع الدخل (كما في شكل (2.11)) حيث أصبحت للناس خيارات عديدة لسلوكهم الادخاري. بالمثل عمومًا الشركات ذات الأرباح العالية يكون من المتوقع أن يكون لديها تنوع كثير في سياسات التقسيم أكبر من الشركات ذات الأرباح المنخفضة. أيضًا الشركات التي مازالت تحت الإنشاء يكون لها تباين أكبر في تقسيم نسب الدخل عن الشركات المنشأة بالفعل.

---

(1) انظر Stefan Valavanis, Econometrics, McGraw-Hill, New York, 1959, p. 48.

(2) As Valavanis puts it, "Income grows, and people now barely discern dollars whereas previously they discerned dimes," ibid., p. 48.



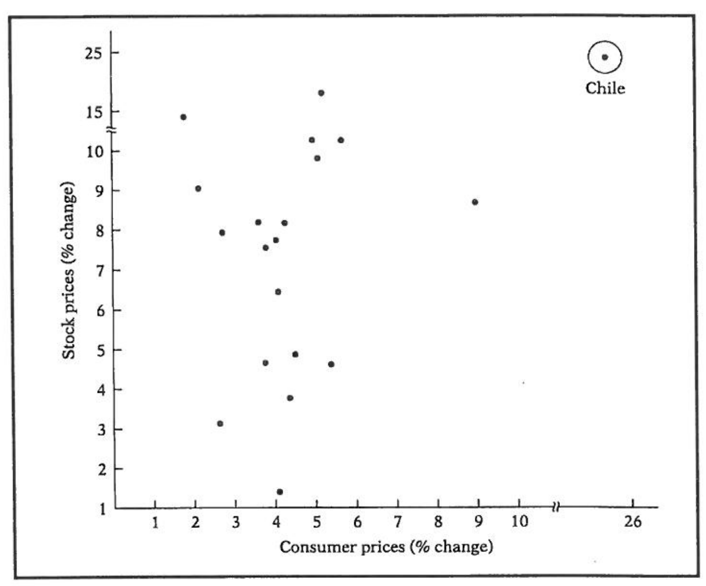
شكل (3.11) يربط بين عدد الأخطاء المطبعية في زمن محدد

3 - كلما زادت جودة أسلوب جمع البيانات، كلما قلت σ_i^2. وبالتالي فإن بعض البنوك يكون لديها أساليب فنية خاصة بالبيانات المعقدة، ويفترض أن تكون الأخطاء فيها قليلة سواء في كشوف الحسابات الخاصة بالعملاء الشهرية أو الربع سنوية، وذلك يكون غير متحقق في البنوك الأخرى التي ليست لديها مثل هذه التقنيات.

4 - اختلاف التباين قد يحدث أيضًا بسبب وجود قيم شاذة في البيانات. فالمشاهدة الشاذة، والتي تعتبر مشاهدة مختلفة بشكل ملحوظ عن باقي المشاهدات (سواء صغير جدًا أو كبيرة جدًا). بشكل أكثر دقة، المشاهدة الشاذة هي مشاهدة من مجتمع مختلف عن الذي تم توليد منه باقي مشاهدات العينة.[3] وجود أو حذف مثل هذه المشاهدة من البيانات خصوصًا إذا كان حجم العينة صغيرًا من الممكن أن يؤثر بشكل كبير على نتائج تحليل الانحدار.

لتوضيح ذلك بمثال، دعنا نعتبر الشكل البياني الموجود في الشكل (4.11) بناء على البيانات الموجودة في تمرين (22.11). يمثل الشكل البياني معدل التغبير في أسعار الأسهم (Y) وسعر المستهلك (X) أثناء فترة الحرب العالمية الثانية لـ 20 دولة. في هذا الشكل البياني مشاهدة Y و X لدولة تشيلي يمكن النظر إليها كمشاهدة شاذة، حيث إن القيمة المعطاة لـ Y و X تعتبر كبيرة جدًا مقارنةً بباقي الدول. في مثل هذا الموقف، يكون من الصعب تحقق فرض ثبات التباين. في تمرين (22.11) يسأل القارئ عن التغيير الذي سيطرأ على نتائج الانحدار إذا تم حذف مشاهدة تشيلي من التحليل.

(3) أشكر Michael McAleer الذي أوضح لي هذه النقطة .



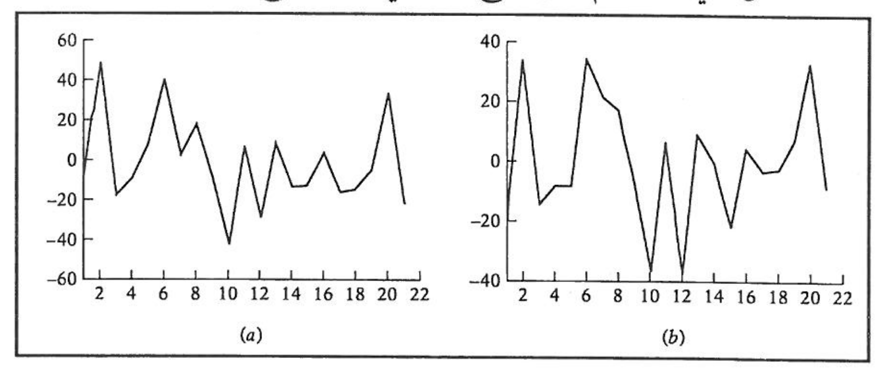
شكل (4.11) العلاقة بين أسعار الأسهم وسعر المستهلك

5 – مصدر آخر لاختلاف التباين، هو عدم تحقق الفرض 9 في CLRM والذي يفترض أن نموذج الانحدار موصف بشكل سليم. وعلى الرغم من أننا سنناقش موضوع أخطاء التوصيف بشكل أكثر توسعًا في (الفصل 13)، إلا أننا يجب أن نعرف الآن أن العديد من اختلاف التباين يعود إلى حقيقة حذف متغيرات مهمة من النموذج. فمثلاً في دالة الطلب على سلعة ما، إذا لم يشتمل النموذج على سعر السلعة المكملة أو السلعة المنافسة للسلعة محل الدراسة (تحيز المتغير المحذوف)، فإن البواقي التي سيتم الحصول عليها من الانحدار قد تعطي الإيحاء بأن تباين الأخطاء غير ثابت، ولكن إذا اشتمل النموذج على هذه المتغيرات المحذوفة، فإن هذا الانطباع قد يزول تمامًا.

كمثال تطبيقي، دعنا نعود إلى الدراسة الخاصة بعائد الإعلان (Y) وعلاقته بتكلفة الإعلان (X). [انظر تمرين (32.8)]. إذا قمنا بعمل انحدار Y على X فقط ولاحظنا بواقي الانحدار، سنرى نمطًا واحدًا، ولكن إذا قمنا بعمل انحدار لـ Y على X و X^2 سنرى أنماطًا أخرى، وذلك يتضح في الشكل (5.11). وقد سبق ورأينا أن X^2 بالفعل تنتمي إلى النموذج (انظر تمرين 32.8).

6 - مصدر آخر لاختلاف التباين هو التواء توزيع واحد وأكثر من المتغيرات المنحدرة الموجودة في النموذج. أمثلة على ذلك، متغيرات اقتصادية مثل الدخل، الثروة، والتعليم. من المعروف جيداً أن توزيع الدخل والثورة في العديد من المجتمعات غير متساو مع وجود معظم الدخل والثروة في يد القليل من الأشخاص الموجودين في قمة المجتمع.

7 - مصادر أخرى لاختلاف التباين: David Hendry حدد أسبابًا أخرى لاختلاف التباين وهي: (1) تحويل خاطئ للبيانات (مثل تحويلة النسبة أو تحويلة الفروق الأولى) و(2) شكل دالي غير سليم (مثل نماذج خطية في مقابل نماذج خطية - لوغاريتمية).[4]



شكل (5.11) بواقي انحدار (a) أي الإعلان ونفقاته و(b) أثر Adexp و Adexp2

لاحظ أن مشكلة اختلاف التباين توجد بشكل أكبر في البيانات المقطعية عن السلاسل الزمنية. في البيانات المقطعية، فإن الفرد عادة ما يتعامل مع مشاهدات من المجتمع عند نقطة زمنية محددة مثل المستهلكين أو أسرهم، مصانع أو مناطق جغرافية مثل ولاية، بلد، مدينة، وإلى ما غير ذلك. بالإضافة إلى أن هذه المشاهدات قد تكون مختلفة في الحجم مثل مصانع صغيرة، متوسطة أو كبيرة أو دخل محدود ومتوسط وكبير. على العكس في بيانات السلاسل الزمنية، فإن المتغيرات عادة ما يكون لها قوى متساوية، لأن الفرد عادة ما يجمع بيانات عن نفس الوحدة خلال فترة زمنية معينة. أمثلة على ذلك GNP، نفقات الاستهلاك والادخار أو العمالة في الولايات المتحدة الأمريكية، مثلاً خلال الفترة 1950 إلى 2000.

لتوضيح وجود اختلاف التباين في تحليل البيانات المقطعية، دعنا نعود إلى جدول (1.11). هذا الجدول يعطي بيانات عن تعويضات العاملين في 10 مصانع

(4) David F. Hendry, Dynamic Econometrics, Oxford University Press, p. 45.

لصناعة السلع غير المعمرة، هذه البيانات مقسمة وفقًا لحجم العمالة في المصنع لعام 1958 . ومعطى أيضًا في الجدول متوسط الإنتاجية لتسع طبقات لحجم العمالة .

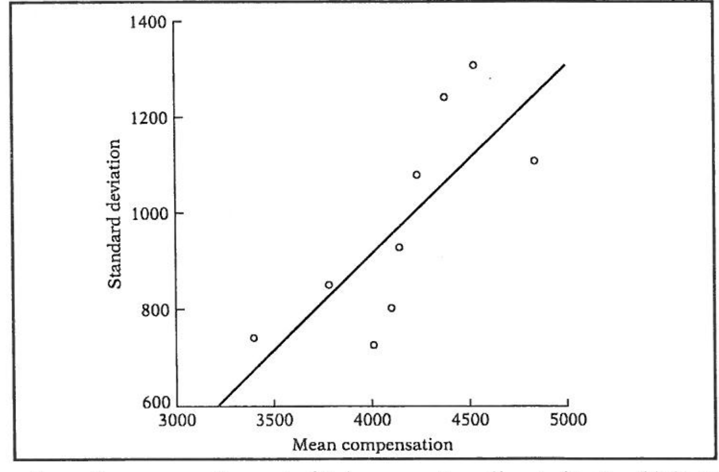
على الرغم من أن الصناعات تختلف في تكوين الناتج الخاص بها، إلا أن جدول (1.11) يوضح أنه في المتوسط، فإن المصنع الكبير يدفع أكثر من المصنع الصغير .

جدول (1.11)

التعويض لكل عامل ($) في صناعات خاصة بالسلع غير المعمرة بناء على حجم العمالة بالمصنع ، 1958

	Employment size (average number of employees)								
Industry	1–4	5–9	10–19	20–49	50–99	100–249	250–499	500–999	1000–2499
Food and kindred products	2994	3295	3565	3907	4189	4486	4676	4968	5342
Tobacco products	1721	2057	3336	3320	2980	2848	3072	2969	3822
Textile mill products	3600	3657	3674	3437	3340	3334	3225	3163	3168
Apparel and related products	3494	3787	3533	3215	3030	2834	2750	2967	3453
Paper and allied products	3498	3847	3913	4135	4445	4885	5132	5342	5326
Printing and publishing	3611	4206	4695	5083	5301	5269	5182	5395	5552
Chemicals and allied products	3875	4660	4930	5005	5114	5248	5630	5870	5876
Petroleum and coal products	4616	5181	5317	5337	5421	5710	6316	6455	6347
Rubber and plastic products	3538	3984	4014	4287	4221	4539	4721	4905	5481
Leather and leather products	3016	3196	3149	3317	3414	3254	3177	3346	4067
Average compensation	3396	3787	4013	4104	4146	4241	4388	4538	4843
Standard deviation	742.2	851.4	727.8	805.06	929.9	1080.6	1241.2	1307.7	1110.5
Average productivity	9355	8584	7962	8275	8389	9418	9795	10,281	11,750

المصدر : تعداد المصانع ، U.S ، قطاع التجارة ، 1958 (محسوبة عن طريق المؤلف)



شكل (2.11) الانحراف المعياري للتعويض ومتوسطه الحسابي **Stardard denation of compencation**

وكمثال دعنا نفترض أن هناك مصنعًا يعين من عامل إلى أربعة عمال، يدفع في المتوسط حوالي 3396 $، في حين أن المصانع التي تعين من 1000 إلى 2499 عامل تدفع في المتوسط 4843 $. ولكن لاحظ أنه يوجد تنوع ملحوظ في المكسب وفقًا لطبقات العمال المختلفة كما هو موضح بالانحراف المعياري المقدر للمكسب. يمكن ملاحظة ذلك أيضًا من خلال شكل (6.11) والذي يوضح العلاقة بين الانحراف المعياري للتعويض، ومتوسط التعويض لكل طبقة من طبقات العمال. من الواضح أنه في المتوسط فإن الانحراف المعياري للتعويض يزداد مع زيادة متوسط قيمة التعويض.

2.11 تقديرات OLS في وجود اختلاف التباين: OLS ESTIMATION IN THE PRESENCE OF HETEROSCEDASTICITY

ما الذي سيحدث لمقدرات OLS وتبايناتها إذا كان لدينا اختلاف في التباين، أي إذا كان $E(u_i^2) = \sigma_i^2$ مع افتراض بقاء كل فروض النموذج التقليدي الأخرى متحققة؟ للإجابة على هذا السؤال. دعنا نسترجع النموذج ثنائي المتغيرات التالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

بتطبيق المعادلة التقليدية، مقدر OLS لـ β_2 هو

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

$$= \frac{n\sum X_i Y_i - \sum X_i \sum Y_i}{n\sum X_i^2 - (\sum X_i)^2} \qquad (1.2.11)$$

ولكن تباين $\hat{\beta}_2$ سنحصل عليه الآن من المعادلة التالية (انظر ملحق A 11 والفقرة 1.A11)

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2} \qquad (2.2.11)$$

والذي يختلف بشكل واضح عن معادلة التباين العادية التي نحصل عليها بافتراض تحقق فرض ثبات التباين والتي يمكن التعبير عنها كالتالي:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad (3.2.11)$$

بالطبع إذا كانت $\sigma_i^2 = \sigma^2$ لكل i، فإن المعادلتين متماثلتان (لماذا؟) تذكر أن $\hat{\beta}_2$ هو أفضل مقدر خطي غير متحيز (BLUE) إذا افترضنا تحقق فروض النموذج

التقليدي، بما فيها فرض ثبات التباين. هل سيظل المقدر BLUE إذا أسقطنا فقط فرض ثبات التباين وبدلناه بفرض اختلاف التباين؟ من السهل إثبات أن $\hat{\beta}_2$ مازال خطيًا وغير متحيز. في الحقيقة، كما يتضح من ملحق A3، الفقرة 2.A3، فإن تحقق عدم التحيز لـ $\hat{\beta}_2$ لا يفترض بالضرورة ثبات التباين للخطأ (u_i). في الحقيقة، تباين u_i ثابت أو متغير، ليس له أي دور في تحديد خاصية عدم التحيز.

تذكر أنه في ملحق A3، الفقرة 7.A3، أثبتنا أن $\hat{\beta}_2$ مقدر متسق تحت صحة فروض نموذج الانحدار الخطي التقليدي. ورغم أننا لن نقوم بإثبات العبارة التالية، إلا أنه من الممكن إثباتها وهي أن $\hat{\beta}_2$ مقدر متسق بغض النظر عن اختلاف التباين، أي أنه كلما زاد حجم العينة فإن β_2 المقدرة تقترب من القيمة الحقيقية بغض النظر عن اختلاف التباين.

ومن الممكن أيضًا إثبات أنه تحت فروض معينة (تسمى فروض الاعتيادية) فإن $\hat{\beta}_2$ لها توزيع معتاد تقريبي. بالطبع كل ما سبق وذكرناه عن $\hat{\beta}_2$ متحقق لكل المعاملات الأخرى الموجودة في نموذج الانحدار المتعدد.

من المؤكد الآن أن $\hat{\beta}_2$ مازال مقدرًا خطيًا غير متحيز ومتسق. هل هو مقدر كاف أو أفضل مقدر أم لا؟ أي هل له أقل تباين داخل فئة المقدرات غير المتحيزة؟ وهل هذا التباين الأقل معطى في المعادلة (2.2.11)؟ إجابة السؤالين هي لا: $\hat{\beta}_2$ لم يعد أفضل مقدر وليس له التباين الأقل المعطى في (2.2.11). إذن ماذا عن الـ BLUE في حالة وجود اختلاف في التباين؟ الإجابة معطاة في الفقرة التالية.

3.11 طريقة المربعات الصغرى العامة (GLS):
THE METHOD OF GENERALIZED LEAST SQUARES (GLS)

لماذا لم يعد مقدر OLS العادي لـ β_2 والمعطى في (1.2.11) الأفضل، على الرغم من أنه غير متحيز؟ يمكننا أن نفهم السبب لذلك من جدول (1.11) فكما يوضح الجدول، هناك تنوع في مكاسب الطبقات المختلفة من العمال. فإذا قمنا بعمل انحدار للتعويض على حجم العمالة، فإننا على الأرجح سنستخدم معرفتنا بأن هناك تباينًا كبيرًا للمكسب بين الطبقات المختلفة.

بالطبع نحن نريد طريقة تقدير، بحيث تعطي المشاهدة المنتمية لمجتمع تباينه كبير وزن أقل عن تلك المنتمية لمجتمع تباينه أقل. وفقًا لجدول (1.11)، فإننا نريد إعطاء

أوزان للمشاهدات الخاصة بالعمال من الطبقات 19–10 و 49–20 أكبر من تلك المعطاة إلى العمال من الطبقات 9–5 و 499–250 حيث إنها في المجمعة الأخيرة التمركز حول الوسط الحسابي أكبر من المتحقق في المجموعة الأولى، مما يسعدنا في تقدير الـPRF بدقة أكثر.

للأسف، طريقة OLS التقليدية لا تتبع مثل هذه المنهجية، وبالتالي لا تستفيد من "المعلومة" الخاصة بعدم تساوي التباين في المتغير التابع Y، مثلاً متغير تعويض العمال الموجود في جدول (1.11): يفترض أوزانًا متساوية أو أهمية متساوية لكل المشاهدات. ولكن في طريقة تقدير أخرى معروفة باسم المربعات الصغرى العامة (GLS) فإننا نضع المعلومات الأخرى في الاعتبار، وبالتالي نحصل على مقدرات BLUE. لفهم هذه الطريقة أكثر دعنا نستخدم نموذج الانحدار المعتاد ذا المتغيرين.

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (1.3.11)$$

وللتسهيل الجبري يمكن كتابته كالتالي:

$$Y_i = \beta_1 X_{0i} + \beta_2 X_i + u_i \qquad (2.3.11)$$

حيث $X_{0i} = 1$ لكل i. يمكن للقارئ أن يتأكد من أن المعادلتين السابقتين متطابقتان:

الآن افترض أن هناك اختلافًا في التباين و σ_i^2 معروفة. بقسمه (2.3.11) على σ_i نحصل على:

$$\frac{Y_i}{\sigma_i} = \beta_1 \left(\frac{X_{0i}}{\sigma_i}\right) + \beta_2 \left(\frac{X_i}{\sigma_i}\right) + \left(\frac{u_i}{\sigma_i}\right) \qquad (3.3.11)$$

لتسهيل العرض يمكن كتابتها كالتالي:

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + u_i^* \qquad (4.3.11)$$

حيث إن المتغير المحول (*) يساوي المتغير الأصلي مقسوم على σ_i (المعروفة). تستخدم β_1^* و β_2^* للتعبير عن معاملات النموذج المحول. وللتفرقة بينهما وبين β_1 و β_2 الخاصة بمعاملات OLS التقليدية. ما هو الهدف من هذه التحويلة؟ للإجابة على ذلك لاحظ الخصائص التالية لمقدار الخطأ المحول u_i^*:

$$\begin{aligned}\operatorname{var}(u_i^*) = E(u_i^*)^2 &= E\left(\frac{u_i}{\sigma_i}\right)^2 \\ &= \frac{1}{\sigma_i^2}E(u_i^2) \qquad \text{since } \sigma_i^2 \text{ is known} \\ &= \frac{1}{\sigma_i^2}(\sigma_i^2) \qquad \text{since } E(u_i^2) = \sigma_i^2 \\ &= 1\end{aligned} \tag{5.3.11}$$

لاحظ أنه مقدار ثابت . أي أن تباين مقدار الخطأ المحول u_i^* يتميز الآن بثبات التباين . وبما أننا مازلنا نفترض صحة باقي فروض النموذج التقليدي، بالإضافة إلى أن u^* الآن يتميز بثبات التباين، بالتالي فإنه إذا طبقنا OLS على النموذج المحول (3.3.11) سنحصل على مقدرات BLUE .

أي للاختصار فإن β_1^* وβ_2^* المقدرة الآن المقدرات BLUE وليست مقدرات OLS العادية لـ $\hat{\beta}_1$، $\hat{\beta}_2$.

هذه الطريقة الخاصة بتحويل المتغيرات الأصلية - كما سبق واستعرضناها - تعتبر تحويل متغيرات لاستيفاء شروط النموذج التقليدية، وبالتالي تطبيق OLS أصبح معروفًا باسم المربعات الصغرى العامة (GLS) أي للاختصار فإن GLS هي OLS للمتغيرات المحولة المستوفية لشروط المربعات الصغرى التقليدية . المقدرات التي حصلنا عليها تسمى مقدرات GLS، وهذه المقدرات هي المقدرات الـ BLUE الطريقة الفعلية للحصول على التقديرات β_1^* وβ_2^* هي كالتالي :

أولاً، نكتب SRF للمعادلة (3.3.11)

$$\frac{Y_i}{\sigma_i} = \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) + \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right) + \left(\frac{\hat{u}_i}{\sigma_i}\right)$$

أو

$$Y_i^* = \hat{\beta}_1^* X_{0i}^* + \hat{\beta}_2^* X_i^* + \hat{u}_i^* \tag{6.3.11}$$

الآن نحصل على مقدرات GLS، ونصغر المقدار

$$\sum \hat{u}_i^{2*} = \sum (Y_i^* - \hat{\beta}_1^* X_{0i}^* - \hat{\beta}_2^* X_i^*)^2$$

أي

$$\sum \left(\frac{\hat{u}_i}{\sigma_i}\right)^2 = \sum \left[\left(\frac{Y_i}{\sigma_i}\right) - \hat{\beta}_1^*\left(\frac{X_{0i}}{\sigma_i}\right) - \hat{\beta}_2^*\left(\frac{X_i}{\sigma_i}\right)\right]^2 \tag{7.3.11}$$

الطريقة الفعلية لتصغير (7.3.11) تتيح نفس الطريقة الحسابية التقليدية المعطاة في الملحق A11، الفقرة 2.A11. كما هو موضح هناك فإن مقدر GLS لـ β_2^* هو

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \tag{8.3.11}$$

وتباينه

$$\operatorname{var}(\hat{\beta}_2^*) = \frac{\sum w_i}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \tag{9.3.11}$$

حيث $w_i = \frac{1}{\sigma_i^2}$.

الفرق بين OLS و GLS : Difference between OLS and GLS

بالعودة إلى (الفصل 3) رأينا أن OLS تصغر المقدار التالي :

$$\sum \hat{u}_i^2 = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2 \tag{10.3.11}$$

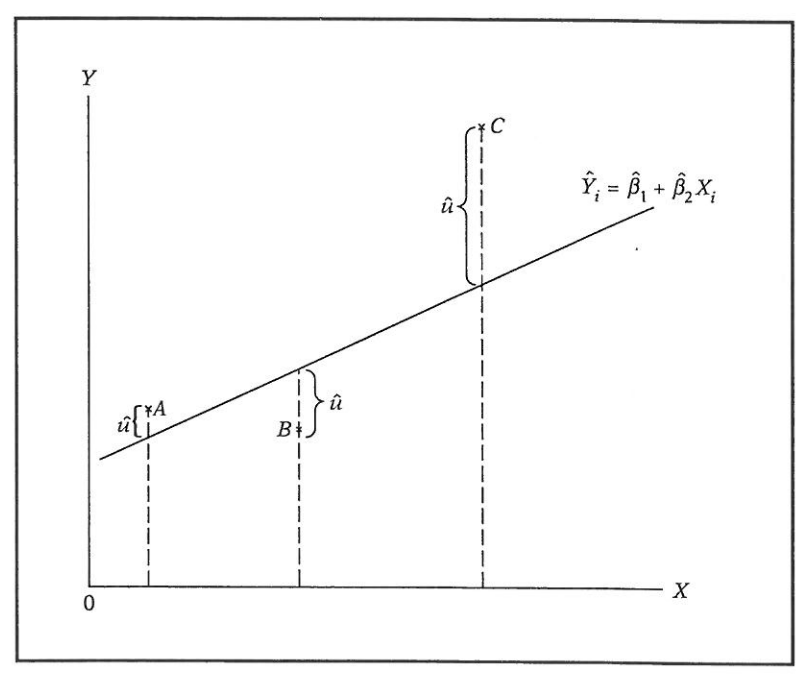
ولكن في الـ GLS نحن نصغر المقدار (7.3.11) والذي يمكن كتابته كالتالي :

$$\sum w_i \hat{u}_i^2 = \sum w_i (Y_i - \hat{\beta}_1^* X_{0i} - \hat{\beta}_2^* X_i)^2 \tag{11.3.11}$$

حيث $w_i = \frac{1}{\sigma_i^2}$ [أثبت أن (11.3.11) و(7.3.11) متطابقين] وبالتالي في GLS نحن نصغر مجموع مربعات البواقي المرجح بالمقدار $w_i = \frac{1}{\sigma_i^2}$ والذي يظهر كأوزان للترجيح، ولكن في OLS نحن نصغر RSS غير مرجح أو مرجح بأوزان متساوية (نفس الوزن لكل المفردات) كما يوضح (7.3.11) فإنه في GLS الأوزان المعطاة لكل مفرد تناسب عكسيًا مع تباينها σ_i، أي أن المشاهدات الآتية من مجتمع تباينه أكبر σ_i سيكون لها وزن أقل نسبيًا من المشاهدات الآتية من مجتمعات تباينها أقل (σ_i) وبالتالي سيكون لهذه المشاهدات الأخيرة وزن أكبر في تصغير الـ RSS (11.3.11) . لمشاهدة الفرق بين OLS و GLS بوضوح دعنا نفترض الشكل البياني الافتراضي الموجود في شكل (7.11) .

في OLS (غير المرجحة)، كل $\hat{u}_i^2$ مرتبطة بالنقاط A، B، C سنحصل على نفس الوزن في تصغير RSS. من الواضح أن المرتبط بالنقطة C سيكون له وزن أقل من المشاهدتين الأخريين، وكما سبق وذكرنا، هذا هو الأسلوب الأفضل في تقدير دالة

انحدار المجتمع (PRF) بشكل أكثر واقعية. فنحن نريد أن نعطي وزنًا أكبر للمشاهدة الأكثر تلاصقًا أو قربًا من المتوسط (المجتمع) عن المشاهدات التي تبتعد وتنتشر بشكل شاذ عن باقي المشاهدات.



شكل (7.11) شكل انتشار بياني افتراضي **Hypothetical scattergram**

بما أن (11.3.11) تصغر RSS المرجح، فإنه من المتوقع أن تعرف باسم المربعات الصغرى المرجحة (WLS) والمقدرات التي نحصل عليها من (8.3.11) و(9.3.11) تعرف باسم المقدرات المرجحة WLS. ولكن WLS حالة خاصة من أسلوب التقدير العام، GLS. في إطار ثبات التباين، من الممكن معاملة WLS وGLS بشكل تبادلي. في فصول قادمة، سنستعرض حالات خاصة أخرى من الـ GLS.

وكملاحظة عابرة نجد أنه إذا كان $w_i = w$، أي ثابت لكل i، فإن β_2^* يتطابق مع $\hat{\beta}_2$ ويكون Var (β_2^*) متطابقًا مع تباين ($\hat{\beta}_2$) العادي (ثبات التباين) المعطى في (3.2.11)، وتلك الملحوظة يجب ألا تكون مفاجأة للقارئ (لماذا؟) (انظر تمرين 8.11).

4.11 عواقب استخدام OLS في حالة وجود اختلاف في التباين: CONSEQUENCES OF USING OLS IN THE PRESENCE OF HETEROSCEDASTICITY

كما سبق ورأينا، فإن كلاً من β_2^* و $\hat{\beta}_2$ مقدرات (خطية) غير متحيزة!. أي أنه في العينات المتكررة، في المتوسط، β_2^* و $\hat{\beta}_2$ سيساوي كل منهما القيمة الحقيقية لـ β_2، أي أن له أقل تباين. ما الذي سيحدث إلى فترة الثقة، اختبارات الفروض وباقي الطرق إذا استخدمنا تقدير OLS إلى β_2^* ؟ سنفرق بين حالتين:

تقدير OLS مع السماح باختلاف التباين:
OLS Estimation Allowing for Heteroscedasticity

افترض أننا نستخدم $\hat{\beta}_2$ والتباين المعطى في المعادلة (2.2.11)، أي الذي يأخذ في اعتباره بوضوح اختلاف التباين. باستخدام هذا التباين وبافتراض أن σ_i^2 . معلومة، هل تستطيع تكوين فترات الثقة، واختبارات الفروض باستخدام اختبارات t و F المعتادة؟ الإجابة العامة لهذا السؤال هي لا، لأنه من الممكن إثبات أن $\text{var}(\beta_2^*) \leq \text{var}(\hat{\beta}_2)$ [5] وذلك يعني أن فترات الثقة - بناء على ما سبق - لن تكون كبيرة بالضرورة. وكنتيجة لذلك فإن اختبارات t و F ستعطي نتائج غير دقيقة، حيث إن $\text{var}(\hat{\beta}_2)$ سيكون كبيرًا والذي سيحدث هو التوصل إلى عدم معنوية المعامل (حيث إن قيمة t ستكون أصغر من المفروض) الذي قد يكون معنويًا بالفعل في الحقيقة، وهذه النتيجة السليمة يمكن أن نتوصل إليها إذا كونا فترات الثقة السليمة بناء على طريقة GLS.

تقدير OLS مع اهمال اختلاف التباين:
OLS Estimation Disregarding for Heteroscedasticity

سيصبح الموقف أكثر خطورة ليس فقط باستمرار استخدام $\hat{\beta}_2$ ولكن أيضًا مع استمرار استخدام معادلة التباين (ثبات التباين) العادية المعطاة في (3.2.11) حتى وإن كان اختلاف التباين موجودًا أو مشكوكا في وجوده: لاحظ أن هذه الحالة هي الأكثر حدوثًا، حيث إن استخدام انحدار OLS التقليدي عن طريق أي حزمة إحصائية

(5) الإثبات الكامل موجود في Phoebus J. Dhrymes, Introductory Econometrics, Springer-Verlag, New york, 1978, pp. 110-111.

ولاحظ إن مقدار خسارة الكفاءة لـ $\hat{\beta}_2$ [أي المقدار الذي يزيد به $\text{var}(\hat{\beta}_2)$ عن $\text{var}(\beta_2^*)$] يعتمد على قيم القيمة الخاصة بالمتغيرات X وقيمة σ_i^2 .

وتجاهل اختلاف التباين (أو عدم ملاحظته) سيؤدي إلى استخدام تباين $\hat{\beta}_2$ المعطى في (3.2.11). أولاً، $\text{Var}(\hat{\beta}_2)$ المعطى في (3.2.11) هو مقدر متحيز لـ $\text{Var}(\hat{\beta}_2)$ المعطى في (2.2.11)، أي أنه في المتوسط فإنه يقدر الأخير بأقل أو أعلى من قيمته الحقيقية. وفي العموم لا تستطيع أن تحدد ما إذا كان مقدار التحيز موجبًا (تقدير بأكثر من القيمة الحقيقية) أو سالبًا (أقل من القيمة الحقيقية) حيث إن ذلك يعتمد على طبيعة العلاقة بين σ_i^2 والقيمة الخاصة بالمتغير المفسر X، ويمكن رؤية ذلك بوضوح في (2.2.11) [انظر تمرين (9.11)]. التحيز يظهر من حقيقة أن $\hat{\sigma}^2$، المقدر الخاص بـ σ^2، يساوي $\sum \hat{u}_i^2/(n-2)$ لم يعد مقدراً غير متحيز لـ σ^2 عندما يظهر اختلاف التباين (انظر ملحق 3.A11) وكنتيجة لذلك لا يمكننا الاعتماد على فترات الثقة المسحوبة بناء عليه أو اختبارات t، F.[6]

للاختصار، إذا صممنا على استخدام طرق الاختبار العادية بغض النظر عن اختلاف التباين، فإن الاستنتاجات التي سنحصل عليها والاستدلال الذي سنتوصل إليه يكون غير سليم وخاطئ.

لإلقاء المزيد من الضوء على هذه النقطة، سنشير إلى دراسة Monte Carlo قام بها Davidson و Mackinnor.[7] فقد اعتبرا النموذج البسيط التالي، الذي نرمز له برموز خاصة بنا كالتالي:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (1.4.11)$$

وقد افترضنا أن $\beta_1 + 1$، $\beta_2 = 1$ و $u_i - N(0, X_i^{\alpha})$. وكما يتضح من الافتراض الأخير، فإن تباين الخطأ غير ثابت ومرتبط بقيمة المتغير المنحدر X مرفوع للأس α. أي أنه على سبيل المثال إذا كانت $\alpha = 1$ فإن تباين الخطأ يتناسب مع قيمة X، أما إذا كانت $\alpha = 2$ فإن تباين الخطأ يتناسب مع مربع قيمة X وهكذا. في الفقرة 6.11 سنتطرق إلى المنطق وراء مثل هذه الطريقة. بناء على 20,000 تكرار، وبالسماح إلى قيم مختلفة من α، حصلوا على الأخطاء القياسية لمعاملي الانحدار باستخدام OLS [انظر المعادلة

(6) من (6.3.5) نعرف ان $100(1-\alpha)\%$ فترة ثقة لـ β_2 هي $\left[\hat{\beta}_2 \pm t_{\alpha/2}\, \text{se}(\hat{\beta}_2)\right]$ ولكن إذا كان $\text{se}(\hat{\beta}_2)$ لا يقدر بشكل غير متحيز، كيف يمكننا الثقة بفترة الثقة المحسوبة؟

(7) Russell Davidson and James G. MacKinnon, Estimation and inference in Econometrics, Oxford University Press, New York, 1993, pp. 549–550.

(3.2.11)] و OLS التي تسمح باختلاف التباين [انظر المعادلة (2.2.11)] و GLS [انظر المعادلة (9.3.11)] . تم تلخيص النتائج لبعض القيم المختارة للـ α كالتالي :

Value of α	Standard error of $\hat{\beta}_1$			Standard error of $\hat{\beta}_2$		
	OLS	OLS_{het}	GLS	OLS	OLS_{het}	GLS
0.5	0.164	0.134	0.110	0.285	0.277	0.243
1.0	0.142	0.101	0.048	0.246	0.247	0.173
2.0	0.116	0.074	0.0073	0.200	0.220	0.109
3.0	0.100	0.064	0.0013	0.173	0.206	0.056
4.0	0.089	0.059	0.0003	0.154	0.195	0.017

لاحظ أن : OLShet تعني OLS التي تضع في اعتبارها اختلاف التباين

الخاصية الأكثر غرابة من هذه النتائج والخاصة بالـ OLS، أنه سواء وضعنا في الاعتبار اختلاف التباين أو لم نضعه في الاعتبار، فإن OLS تقدر الخطأ المعياري بشكل متسق بقيمة أزيد من قيمته الحقيقية التي سنحصل عليها إذا استخدمنا طريقة GLS (الصحيحة)، خصوصًا لقيمة كبيرة من α، والتي تزيد من جودة تقدير GLS . هذه النتائج توضح أيضًا أنه إذا لم تستخدم GLS واعتمدنا فقط على الـ OLS – سواء وضعنا في الاعتبار اختلاف التباين أم لا – فإن الصورة تكون مختلطة . فالـ OLS العادية لها أخطاء معيارية إما كبيرة جدًا (للجزء المقطوع من المحور الصادي) أو بشكل عام صغير جدًا (لمعامل الميل) مقارنةً مع الـ OLS التي نضع فيها اختلاف التباين في الاعتبار . وبالتالي فالرسالة واضحة : في وجود اختلاف التباين، استخدم GLS . عمومًا ولأسباب عديدة مذكورة تفصيلياً في هذا الفصل لاحقًا، في الواقع العملي ليس من السهل تطبيق GLS . أيضًا كما سنناقش لاحقًا، إذا لم يكن اختلاف التباين شديد الوضوح، لن يكون من السهل تفصيل GLS و WLS على OLS .

من المناقشة السابقة ، يتضح أن اختلاف التباين مشكلة محتملة خطيرة، وعلى الباحث تحديد ما إذا كان يواجه مثل هذه المشكلة أم لا . إذا تعرف الباحث على وجود هذه المشكلة، فلابد أن يتخذ إجراءات إصلاحية مثل استخدام انحدار المربعات العنصري المرجح أو أي طرق أخرى . وقبل البدء في استعراض الطرق المختلفة لتصحيح هذه المشكلة، يجب أولاً أن نعرف ما إذا كانت مشكلة اختلاف التباين موجودة أصلاً أم لا وفقًا للبيانات محل الدراسة . هذه النقطة سيتم استعراضها في الفقرة التالية :

ملحوظة فنية : A Technical Note

على الرغم من أننا سبق وذكرنا، أنه في حالة اختلاف التباين، يجب استخدام GLS وليس OLS، والتي تعتبر BLUE، هناك أمثلة أخرى يمكن أن تكون OLS هي المقدرات الـ BLUE بغض النظر عن اختلاف التباين.[8]

ولكن هذه الأمثلة غير واردة من الناحية التطبيقية.

5.11 اكتشاف اختلاف التباين: DETECTION OF HETEROSCEDASTICITY

كما كان الحال في تعدد العلاقات الخطية، السؤال المهم تطبيقيًا: كيف يمكننا معرفة ما إذا كان اختلاف التباين موجودًا في دراسة معينة؟ مرة أخرى كما في تعدد العلاقات الخطية، لا توجد قواعد سريعة محددة لاكتشاف اختلاف التباين، ولكن هناك بعض المؤشرات العامة. وهذا الموقف محتوم ولا يمكن تجنبه، حيث إن σ_i^2 تكون معروفة فقط إذا عرفت مجتمع الـ y المرتبطة بكل X's مختارة، كما هو الحال في جدول (1.2) أو جدول (1.11) لكن هذه البيانات تعتبر استثناء أكثر منه قاعدة في معظم الدراسات الاقتصادية. وفقًا لهذا المنطلق يختلف باحثو الاقتصاد القياسي عن علماء مجال الزراعة والبيولوجي، حيث إن الباحثين يكون لديهم تحكم جيد على جميع العناصر. في أغلب الأحيان في الدراسات الاقتصادية هناك عينة واحدة للـ Y مرتبطة بكل قيمة معينة لـ X. ولا توجد طريقة لمعرفة σ_i^2 من مفردة وحدة للـ Y. وبالتالي في أغلب الأحيان المرتبطة بدراسات الاقتصاد القياسي، اختلاف التباين يكون هو الوضع القائم نتيجة العينة أو الظن القائم على العلم، تجارب عملية سابقة أو آراء سابقة.

مع وضع هذا الأمر في الاعتبار، دعنا نختبر بعض الطرق الرسمية وغير الرسمية لاكتشاف اختلاف التباين. وكما سنرى في المناقشة التالية، العديد من هذه الطرق يعتمد على اختيار بواقي الـ OLS $\hat{u}_i$ حيث إن قيمهم يمكن مشاهدتها وليس توزيع u_i.

(8) سبب ذلك هو نظرية Gauss-Markov والخاصة بالشروط الكافية (وغير الضرورية) الـ OLS حتى تصبح مقدرات كفء. الشروط الضرورية والكافية للـ OLS حتى تصبح BLUE معطاة في نظرية Kruskal. ولكن هذه النقطة خارج نطاق الدراسة الحالية وأنا شاكر جداً لـ Michael McAleer لمساعدتى في إيضاح هذه النقطة. للمزيد من التفاصيل، انظر

Denzil G. Fiebi, Michael McAleer, and Robert Barels, "Properties o Ordinary Leas Squares Estimators in Regression Models with Nospherical Disturbances," Journal of Econometrics, vol. 54, No. 1-3, Oct.-Dec., 1992, pp. 321-334.

والباحث يتمنى أن يحصل على مقدرات جيدة لـ u_i وهنا الأمل يمكن أن يتحقق بالفعل إذا كان حجم العينة كبيراً نسبياً.

الطرق غير الرسمية : Informal Methods

طبيعة المشكلة : Nature of the Problem

فى أغلب الأحيان تكون طبيعة المسألة محل الدراسة تقترح وجود اختلاف في التباين. فمثلاً في دراسات عن ميزانيات الأسرة التي قام بها كل من Prais وHouthakker، وجدا أن تباين البواقي حول انحدار الاستهلاك على الدخل يتزايد مع الدخل، وبالتالي أي باحث في مسوح مماثلة يتوقع حدوث عدم تساو في التباين الخاص بالخطأ.[9] في واقع الأمر، في البيانات المقطعية والتي تشتمل على مفردات غير متجانسة يكون اختلاف التباين قاعدة أكثر منه توقعاً. وبالتالي في تحليل البيانات المقطعية والذي يشتمل على نفقات الاستثمار وعلاقاتها بالمبيعات ومعدل الفائدة وغيره. اختلاف التباين يكون متوقعاً بوجه عام، سواء كانت الأحجام صغيرة، متوسطة أو كبيرة.

في واقع الأمر، سبق وتناولنا أمثلة خاصة بذلك. (في الفصل 2) استعرضنا العلاقة بين متوسط الأجر بالساعة، وعدد سنوات الدراسة في الولايات المتحدة. في هذا الفصل، سنناقش أيضاً العلاقة بين نفقات الطعام وإجمالي النفقات لعدد 55 أسرة في الهند. [انظر تمرين (16.11)]

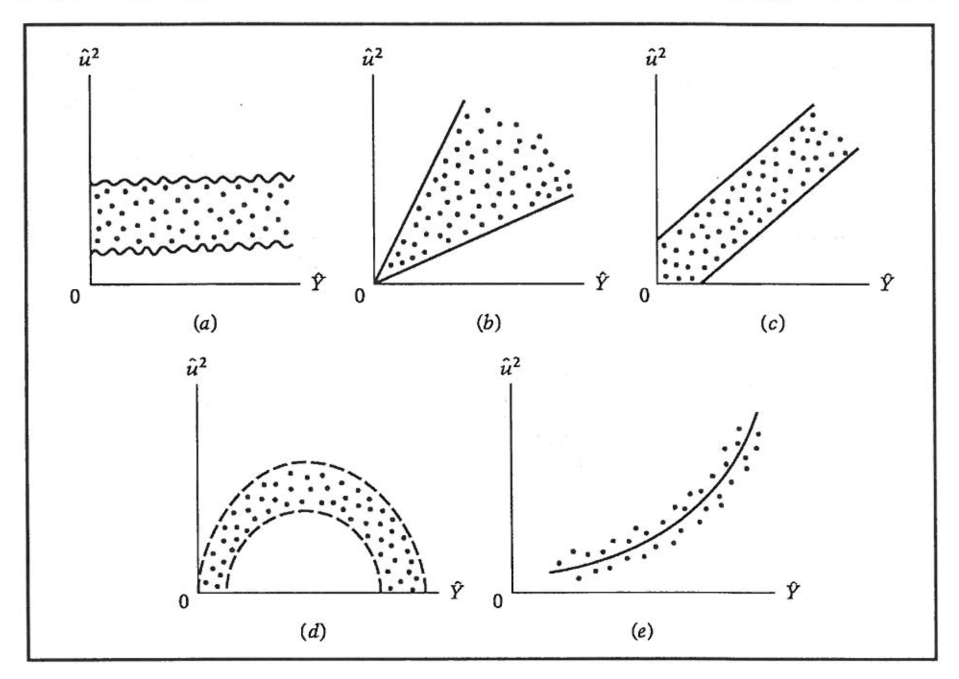
الطريقة البيانية : Graphical Method

إذا لم تكن هناك معلومات سابقة أو فعلية عن طبيعة اختلاف التباين، ففي الواقع يمكن الباحث أن يقوم بعمل تحليل الانحدار بناء على فرض ثبات التباين، وبعد ذلك يقوم بعمل اختبارات لمربعات البواقي $\hat{u}_i^2$ الذي إذا كان هناك أي نمط منتظم لها . وعلى الرغم من أن $\hat{u}_i^2$ ليست لها نفس قيم u_i^2، إلا أنه يستخدم كبديل له خصوصاً إذا كان حجم العينة كبيراً بدرجة كافية.[10] اختبار $\hat{u}_i^2$ قد ينتج عنه أنماط مختلفة كالموجودة في الشكل (8.11).

---

(9) S. J. Prais and H. S. Houthakker, The Analysis of Family Budgets, Cambridge University Press, New York, 1955.

(10) لمزيد من التفاصيل عن العلاقة بين ûi و ui انظر E. Malinvaud, Statistical Methods of Econometrics, North Holland Publishing Company, Amsterdam, 1970, pp. 88–89.



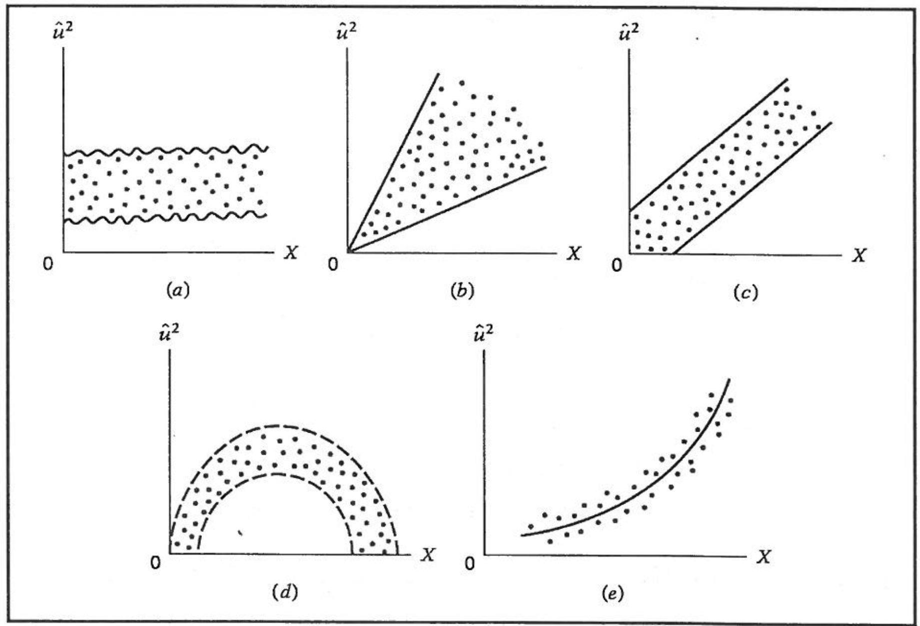
شكل (8.11) أنماط افتراضية لمربعات البواقي المقدرة

Hypothetical patterns of estimated Squared residuals

في الشكل (8.11)، $\hat{u}_i^2$ مرسومة بيانيًا ضد $\hat{Y}_i$ ، قيمة Y_i المقدرة من خط الانحدار. الفكرة وراء ذلك هي تحديد ما إذا كانت متوسط القيم المقدرة للـ Y مرتبطًا بشكل منتظم مع البواقي المربعة أم لا. في شكل (a8.11) نرى أنه لا يوجد أي نمط منتظم بين المتغيرات، مما يعني عدم وجود اختلاف في التباين في البيانات محل الدراسة. في شكل (b8.11 إلى e) نجد أنماطًا مختلفة. فمثلاً، شكل (c8.11) يقترح وجود علاقة خطية، في حين شكل (d8.11 و e) يقترح وجود علاقة تربيعية بين $\hat{u}_i^2$ و $\hat{Y}_i$. وفقًا لهذه المعلومة، ممكن أن يحول الباحث هذه البيانات بطريقة تجعل البيانات المحولة لا يوجد فيها اختلاف التباين. في الفقرة (6.11) سنختبر العديد من هذه التحويلات.

بدلاً من عمل الشكل البياني $\hat{u}_i^2$ ضد $\hat{Y}_i$ ، من الممكن رسم أي منهما ضد واحد من المتغيرات المفسرة، خصوصًا إذا كان الشكل البياني لـ $\hat{u}_i^2$ ضد $\hat{Y}_i$ يظهر كما في الشكل (a8.11). فهذا الشكل، والموضح في شكل (9.11)، قد يشتمل على أنماط مناظرة للموجودة في شكل (8.11) (في حالة وجود نموذج ذي متغيرين اثنين،

رسم $\hat{u}_i^2$ ضد $\hat{Y}_i$ مساوية لرسمه ضد X_i، ولهذا فإن الشكل (9.11) مكافيء لشكل (8.11). ولكن هذا ليس الحال دائمًا إذا اعتبرنا أن هناك نموذجًا يشتمل على أكثر من متغيرين للـX. ففي هذه الحالة يمكن رسم $\hat{u}_i^2$ ضد أي من متغيرات الـX الموجودة في النموذج).



شكل (9.11) شكل انتشار لمربعات البواقي المقدرة ضد X
Scattergram of estimated squared residuals against X

فمثلاً نمط كالموجود في شكل (c9.11) يقترح أن تباين الخطأ مرتبط خطيًا مع المتغير X. وبالتالي إذا كانا في انحدار الادخار على الدخل، فإن الباحث قد يصل إلى نمط كالموجود في شكل (c9.11) وهذا النمط يقترح أن اختلافُ التباين قد يكون متناسبًا مع قيمة متغير الدخل. هذه المعلومة قد تساعدنا في إيجاد تحويلة للبيانات، بحيث يساعد ذلك عند القيام بانحدار البيانات المحولة أن يكون تباينها ثابتًا. سنعود مرة أخرى إلى هذا الموضوع في الفقرة التالية:

(11) R. E. Park, "Estimation with Heteroscedastic Error Terms," Econometrica, vol. 34, no. 4, october 1966, p. 888.

اختيار Park حالة خاصة من الاختيار العام الذي اقترحه A. C. Harvey in "Estimating Regression Models with Multiplicative Heteroscedasticity," Econometrica, vol. 44, no. 3, 1976, pp. 461–465.

الطرق الرسمية : Formal Metdods

اختبار Park [11] :

وضع Park الطريقة البيانية في صورة طريقة رسمية لاكتشاف اختلاف التباين، فافترض أن σ_i^2 يعتبر دالة ما في المتغير المفسر X_i، الشكل الدالي الذي اقترحه هو:

$$\sigma_i^2 = \sigma^2 X_i^{\beta} e^{vi}$$

أو

$$\ln \sigma_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i \qquad (1.5.11)$$

حيث v_i هو متغير الخطأ العشوائي.

بما أن σ_i^2 عموماً غير معلومة، فإن Park اقترح استخدام $\hat{u}_i^2$ كبديل، وعمل الانحدار التالي:

$$\ln \hat{u}_i^2 = \ln \sigma^2 + \beta \ln X_i + v_i$$
$$= \alpha + \beta \ln X_i + v_i \qquad (2.5.11)$$

إذا كانت β معنوية إحصائيًا، فإن ذلك يعني وجود اختلاف في التباين في البيانات محل الدراسة. أما إذا كانت غير معنوية فإننا قد نقبل فرض ثبات التباين. اختبار Park بهذه الطريق يتم على مرحلتين، في المرحلة الأولى نطبق انحدار OLS مع تجاهل السؤال حول وجود اختلاف التباين من عدمه. حصل على $\hat{u}_i$ من هذا الانحدار، ثم ندخل في المرحلة الثانية ونطبق انحدار (2.5.11).

وبالإضافة أنه من الناحية العملية هناك بعض المشاكل في تطبيق اختبار Park. فإن Goldfeld و Quandt ناقشا أيضًا أن مقدار الخطأ v_i الموجود في (2.5.11) قد لا يكون مستوفيًا شروط الـ OLS وبالتالي قد يكون هو نفسه غير ثابت التباين.[12] ولكن بخلاف ذلك ووفقًا لطريقة المتغيرات المفسرة ممكن للباحث أن يستخدم اختبار Park.

مثال 1.11

العلاقة بين التعويضات والإنتاجية :

RELATIONSHIP BETWEEN COMPENSATION AND PRODUCTIVITY

لشرح أسلوب Park، دعنا نستخدم البيانات الموجودة في جدول (1.11) ونقوم بعمل الانحدار التالي:

(12) Stephen M. Goldfeld and Richard E. Quandt, Nonlinear Methods in Econometrics, North Holland Publishing Company, Amsterdam, 1972, pp. 93–94.

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

حيث Y = متوسط تعويضات بآلاف الدولارات، X = متوسط الإنتاجية بآلاف الدولارات و i = حجم العمالة i في المصنع، نتائج الانحدار جاءت كالتالي:

$$\hat{Y}_i = 1992.3452 + 0.2329X_i$$
$$\text{se} = (936.4791) \quad (0.0998) \qquad (3.5.11)$$
$$t = (2.1275) \quad (2.333) \qquad R^2 = 0.4375$$

النتائج توضح أن معامل الميل المقدر معنوي عند مستوى المعنوية 5% على أساس اختبار ذي طرف واحد للـ t. المعادلة توضح أنه كلما زادت إنتاجية العمالة مثلاً بدولار واحد فإن متوسط تعويضات العمالة تزايد بحوالي 23 سنتًا.

البواقي التي نحصل عليها من انحدار (3.5.11) نقوم بعمل انحدار لها على X_i كما هو مقترح في المعادلة (2.5.11) وحصلنا على النتائج التالية:

$$\widehat{\ln \hat{u}_i^2} = 35.817 - 2.8099 \ln X_i$$
$$\text{se} = (38.319) \quad (4.216) \qquad (4.5.11)$$
$$t = (0.934) \quad (-0.667) \qquad R^2 = 0.0595$$

يتضح أنه لا توجد علاقة إحصائية معنوية بين المتغيرين. ووفقًا لقاعدة Park فإننا نستنتج أنه لا يوجد اختلاف في التباين في مقدار الخطأ.[13]

اختبار Glejser[14]:

اختبار Glejser مشابه روحيًا لاختبار Park. فبعد أن نحصل على البواقي $\hat{u}_i$ من انحدار الـ OLS، يقترح Glejser أن نقوم بعمل انحدار للقيم المطلقة لـ $\hat{u}_i$ على المتغير X، والذي يفترض أنه مرتبط بدرجة كبيرة مع σ_i^2. في تجاربه، قام Glejser باستخدام أشكال الدوال الآتية:

$$|\hat{u}_i| = \beta_1 + \beta_2 X_i + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \sqrt{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{X_i} + v_i$$

$$|\hat{u}_i| = \beta_1 + \beta_2 \frac{1}{\sqrt{X_i}} + v_i$$

(13) شكل الدالة الذي يقدمه Park مجرد شكل مقترح. أي شكل آخر قد يؤدي لوجود علاقة معنوية. فمثلاً، إذا استخدمنا $\hat{u}_i^2$ بدلاً من $\ln \hat{u}_i^2$ كمتغير تابع قد نحصل على نتيجة مختلفة.

(14) H. Glejser, "A New Test for Heteroscedasticity," Journal of the American Statistican Association, vol. 64, 1969, pp. 316–323.

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

حيث v_i هو مقدار الخطأ.

مرة أخرى، من الناحية الفعلية أو التطبيقية، يمكن للباحث أن يستخدم أسلوب Glejser ولكن Goldfeld و Quandt أشارا إلى أن مقدار الخطأ v_i قد يكون به بعض المشاكل، حيث إن قيمته المتوقعة ليست بصفر، فإنه مرتبط تسلسليًا (انظر الفصل 12) وبالطبع سيكون به اختلاف في التباين.[15]

وهناك صعوبة إضافية في طريقة Glejser هي أن النماذج مثل:

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

و

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

نماذج غير خطية في المعالم، وبالتالي لا يمكن تقديرها بطريقة OLS العادية. Glejser وجد أنه بالنسبة للعينات الكبيرة، فالنماذج الأربعة الأولى تعطي بوجه عام نتائج مرضية لاكتشاف اختلاف التباين. وبشكل تطبيقي، فإن أسلوب Glejser يمكن أن يستخدم للعينات الكبيرة، أما للعينات الصغيرة فيجب أن يستخدم وفقًا لقيود معينة ليكون أداة كمية جيدة للكشف عن اختلاف التباين.

مثال 2.11

العلاقة بين التعويضات والإنتاجية : اختبار Glejser

RELATIONSHIP BETWEEN COMPENSATION AND PRODUCTIVITY: THE GLEJSER TEST

باستكمال مثال (1.11)، القيمة المطلقة للبواقي تم الحصول عليها من انحدار (3.5.11) حيث تم عمل انحدار على متوسط الإنتاجية (X) وحصلنا على النتائج التالية:

$$\widehat{|\hat{u}_i|} = 407.2783 - 0.0203X_i$$
$$se = (633.1621) \quad (0.0675) \quad r^2 = 0.0127 \qquad (5.5.11)$$
$$t = (0.6432) \quad (-0.3012)$$

نرى من هذا الانحدار أنه لا توجد علاقة بين القيمة المطلقة للبواقي والمتغير المنحدر، متوسط الإنتاجية. هذا الاستنتاج يدعم ما حصلنا عليه من قبل باستخدام اختبار Park.

(15) لمزيد من التفاصيل، انظر Goldfeld and Quandt, op. cit., Chap. 3.

اختبار ارتباط الرتب لـ Spearman:

في تمرين (8.3) عرفنا معامل ارتباط الرتب لـ Spearman كالتالي :

$$r_s = 1 - 6\left[\frac{\sum d_i^2}{n(n^2 - 1)}\right] \qquad (6.5.11)$$

حيث d_i = فرق الرتب بين مشاهدات للمفردة i و n = عدد المشاهدات المرتبة.

معامل ارتباط الرتب السابق ممكن أن يستخدم لاكتشاف اختلاف التباين كالتالي :

افترض أن $Y_i = \beta_0 + \beta_1 X_i + u_i$

خطوة 1 : قم بعمل انحدار لـ Y على X واحصل على البواقي $\hat{u}_i$.

خطوة 2 : تجاهل إشارة $\hat{u}_i$ ، أي نحصل على القيم المطلقة $|\hat{u}_i|$ ، رتب كلاً من $|\hat{u}_i|$ و X_i (أو $\hat{Y}_i$) وفقًا لترتيب تصاعدي أو تنازلي واحسب معامل ارتباط الرتب لـ Spearman كما سبق أن استعرضناه.

خطوة 3 : بافتراض أن معامل ارتباط الرتب للمجتمع ρ_s يساوي الصفر و $n > 8$ معنوية r_s من العينة يمكن اختيارها باستخدام اختبار t كالتالي :[16]

$$t = \frac{r_s\sqrt{n-2}}{\sqrt{1-r_s^2}} \qquad (7.5.11)$$

بدرجات حرية = $\text{df} = n - 2$

إذا زادت قيمة t المحسوبة عن قيمة t الحرجة، قد نقبل فرض اختلاف التباين، وبخلاف ذلك نرفض هذا الفرض. إذا اشتمل نموذج الانحدار على أكثر من متغير X فإن r_s يمكن حسابها بين $|\hat{u}_i|$ وكل متغير من المتغيرات X منفرد ويمكن أن تختبر المعنوية الإحصائية باستخدام اختبار t المعطى في المعادلة (7.5.11).

مثال 3.11

توضيح اختبار ارتباط الرتب ILLUSTRATION OF THE RANK CORRELATION TEST

لشرح اختبار ارتباط الرتب، دعنا نعتبر البيانات المعطاة في جدول (2.11). البيانات خاصة بمتوسط العائد السنوي (E ، %) والانحراف المعياري له (% ، σ_i) لكل 10 مشروعات خط رأس مال السوق (CML) لنظرية الأرباح يفترض وجود علاقة خطية بين

(16) انظر G. Udny Yule and M. G. Kendall, An Introduction to the Theory of Statistics, Charles Griffin & Company, London, 1953, p. 455.

العائد المتوقع (E_i) والمخاطرة (مقاسة بالانحراف المعياري، σ) كالتالي :

$$E_i = \beta_1 + \beta_2 \sigma_i$$

باستخدام بيانات جدول (2.11)، تم تقدير النموذج السابق، وتم حساب بواقي النموذج، وبما أن البيانات خاصة بـ 10 مشروعات بأحجام وأهداف مختلفة، فإننا قد نتوقع مسبقًا وجود اختلاف في التباين، لاختبار هذا الفرض، دعنا نطبق اختبار ارتباط بتطبيق المعادلة (6.5.11) نحصل على

$$r_s = 1 - 6\frac{110}{10(100-1)} \quad (8.5.11)$$
$$= 0.3333$$

وبتطبيق اختبار t الموجود في (6.5.11) نحصل على

$$t = \frac{(0.3333)(\sqrt{8})}{\sqrt{1-0.1110}} \quad (9.5.11)$$
$$= 0.9998$$

باستخدام درجات حرية = 8 ، فإن هذه القيمة للـ t ليست معنوية حتى إذا استخدمنا مستوى معنوية 10%، فقيمة p-valre تساوي 0.17. وبالتالي لا يوجد أي دليل على وجود علاقة منتظمة بين المتغير المفسر والقيمة المطلقة للبواقي، مما يعني عدم وجود اختلاف في التباين.

جدول (2.11) اختبار ارتباط الرتب لاختلاف التباين **Rank Conelation Test of Heteroscedasticity**

Name of mutual fund	E_i, average annual return, %	σ_i, standard deviation of annual return, %	$\hat{E}_i^*$	$\|\hat{u}_i\|^\dagger$ residuals, $\|(E_i - \hat{E}_i)\|$	Rank of $\|\hat{u}_i\|$	Rank of σ_i	d, difference between two rankings	d^2
Boston Fund	12.4	12.1	11.37	1.03	9	4	5	25
Delaware Fund	14.4	21.4	15.64	1.24	10	9	1	1
Equity Fund	14.6	18.7	14.40	0.20	4	7	−3	9
Fundamental Investors	16.0	21.7	15.78	0.22	5	10	−5	25
Investors Mutual	11.3	12.5	11.56	0.26	6	5	1	1
Loomis-Sales Mutual Fund	10.0	10.4	10.59	0.59	7	2	5	25
Massachusetts Investors Trust	16.2	20.8	15.37	0.83	8	8	0	0
New England Fund	10.4	10.2	10.50	0.10	3	1	2	4
Putnam Fund of Boston	13.1	16.0	13.16	0.06	2	6	−4	16
Wellington Fund	11.3	12.0	11.33	0.03	1	3	−2	4
Total							0	110

(*) حصلنا عليه من $\hat{E}_i = 5.8194 + 0.4590\ \sigma_i$.

(†) القيمة المطلقة للبواقي.

لاحظ أن: الترتب هو ترتيب تصاعدي للقيم.

اختبار Goldfeld - Quandt [17] :

هذه الطريقة المعروفة يمكن استخدامها إذا كان هناك افتراض بأن التباينات المختلفة σ_i^2 مرتبطة طرديًا مع أحد المتغيرات المفسرة الموجودة في نموذج الانحدار.

(17) Goldfeld and Quandt, op. cit., Chap. 3.

للتبسيط، دعنا نعتبر النموذج التقليدي الذي يشتمل على متغيرين اثنين:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

افترض أن σ_i^2 مرتبطة طرديًا مع X_i كالتالي:

$$\sigma_i^2 = \sigma^2 X_i^2 \qquad (10.5.11)$$

حيث σ^2 ثابت.[18]

الفرض (10.5.11) يقول بأن σ_i^2 تتناسب مع مربع المتغير X. مثل هذا الافتراض نال استحسانًا كبيرًا من كل من Prais و Houthakker في دراستهما الخاصة بميزانية الأسرة [انظر الفقرة (6.11)].

إذا كان استخدام (10.5.11) مقبولاً فإن ذلك يعني أن σ_i^2 سيزداد مع زيادة قيمة X_i، فإذا حدث ذلك فعلاً، فإن اختلاف التباين سيكون موجودًا في النموذج لاختبار ذلك صراحةً، قام كل من Goldfeld و Quandt باقتراح الخطوات التالية:

الخطوة 1. رتب المشاهدات وفقًا لقيم X، ابدأ بأصغر قيمة لـ X.

الخطوة 2. احذف المشاهدات المركزية c، حيث c يتم تحديدها مسبقًا وقسّم الباقي $(n-c)$ مشاهدة إلى مجموعتين، كل منهما بها عدد $\frac{n-c}{2}$ من المشاهدات.

الخطوة 3. قدر انحدارًا خاصًا بالمشاهدات $\frac{n-c}{2}$ الأولى باستخدام OLS وكذلك انحدار خاص بالمشاهدات $\frac{n-c}{2}$ الأخيرة واحصل في كل من الحالتين على مجموع مربعات البواقي RSS_1 و RSS_2 حيث RSS_1 تمثل RSS من نموذج الانحدار الخاص بالقيم الصغرى لـ X_i (أي المجموعة ذات التباين الأقل) و RSS_2 خاصة بنموذج الانحدار المستخدم فيه قيم X_i الكبرى (أي المجموعة ذات التباين الأكبر) كل من هذين الـ RSS لهما درجات حرية كالتالي:

أما $\left(\frac{n-c-2k}{2}\right)$ df أو $\frac{(n-c)}{2} - k$

حيث k تمثل عدد المعالم المقدرة، مشتملة على الجزء الثابت المقطوع من المحور الصادي (لماذا؟). في حالة وجود متغيرين فإن k بالقطع تساوي 2.

(18) هذا مجرد افتراض لهذه الحالة فقط، ففي الحقيقة المطلوب فعليًا هو أن تكون σ_i^2 مرتبطة بشكل منتظم مع X_i.

الخطوة 4. احسب النسبة

$$\lambda = \frac{RSS_2/df}{RSS_1/df} \qquad (11.5.11)$$

إذا افترضنا أن u_i تتبع التوزيع الطبيعي (وعادة ما نفترض ذلك) وإذا كان فرض ثبات التباين متحققًا، فإنه يمكن إثبات أن λ الموجودة في (11.5.11) ستتبع توزيع F ببسط ومقام لهما درجة حرية $(n-c-2k)/2$.

وبالتالي إذا وجد في تطبيق ما أن λ المحسوبة ($=F$) أكبر من F الحرجة عند مستوى المعنوية المحدد، فإننا نرفض الفرض الخاص بثبات التباين، وبالتالي نستنتج أن اختلاف التباين هو الأكثر احتمالاً.

قبل شرح الاختبار، لابد من توضيح نقطة مهمة خاصة بالـ c مشاهدة المركزية: هذه المشاهدات تم حذفها لتوضيح وإبراز الفرق بين مجموعة التباين القليل (أي RSS_1) ومجموعة التباين الكبير (أي RSS_2) ولكن القدرة على ذلك في إنجاح اختبار Goldfeld - Quandt تعتمد على كيفية اختيار C.[19] بالنسبة للنموذج الذي يشتمل على متغيرين اثنين تجربة المحاكاة التي قام بها كل من Goldfeld و Quandt تقترح أن تكون c مساوية لـ 8 اذا كان حجم العينة قريبًا من 30، ويكون مساويًا لـ 16 إذا كان حجم القيمة قريبًا من 60. ولكن Judge et al لاحظ أن $c = 4$ إذا كانت $n = 30$ و $c = 10$ إذا كانت n قريبة من الـ 60 يعتبر حجمًا مناسبًا ووافيًا في التطبيقات العملية.[20]

قبل الانتقال إلى نقطة أخرى، يمكنك ملاحظة أنه إذا كان لدينا أكثر من متغير X واحد في النموذج، فإن ترتيب المشاهدات، وفقًا للخطوة الأولى للاختبار، يمكن أن يتم وفقًا لأي متغير منها. وبالتالي في النموذج $Y_i = \beta_1 + \beta_2X_{2i} + \beta_3X_{3i} + \beta_4X_{4i} + u_i$، يمكن أن نرتب البيانات وفقًا لأي من هذه الـ X's.

إذا كنت لا تعلم مسبقًا أن بعض هذه المتغيرات سيكون مناسبًا استخدامه، يمكن أن تجري الاختبار لكل منها، أو عن طريق اختبار Park، بالترتب لكل X.

---

(19) فنيًا فإن قوة الاختبار تعتمد على كيفية اختيار الـ c. في الإحصاء قوة الاختبار تقاس باحتمال رفض الفرض العدمي وهو خاطئ [أي 1 − امثال (الخطأ من النوع الثاني)] هنا الفرض العدمي يعتمد على أن تباين المجموعتين متساو، أي هناك ثبات في التباين. لمزيد من التفاصيل، انظر M. M. Ali and C. Giaccotto, "A Study of Several New and Existing Tests for Heteroscedasticity in the General Linear Model," Journal of Econometrics, vol. 26, 1984, pp. 355–373.

(20) George G. Judge, R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee, Introduction to the Theory and Practice of Econometrics, John Wiley & Sons, New York, 1982, p. 422.

مثال 4.11

اختبار GOLDFELD - QUANDT

لشرح اختبار Goldfeld - Quandt، دعنا نستعرض بيانات جدول (3.11) والخاصة بإنفاق الاستهلاك وعلاقته بالدخل لعينة قطاعية خاصة بـ 30 أسرة. دعنا نفترض أن هناك علاقة خطية بين نفقات الاستهلاك والدخل، ولكن مع افتراض اختلاف التباين، ودعنا نفترض أيضًا أن طبيعة اختلاف التباين كالمعطاة في (10.5.11) الترتب المطلوب للبيانات. حتى نتمكن من تطبيق الاختبار الموضح في جدول (3.11).

يحذف الـ 4 مشاهدات الأولى والـ 13 مشاهدة الأخيرة وأيضًا مجموعات البواقي الخاصة بها معطاة في الفقرة التالية (الأخطاء المعيارية معطاة بين الأقواس).

الانحدار وفقًا للـ 13 مشاهدة الأولى

$$\hat{Y}_i = 3.4094 + 0.6968X_i$$

$$(8.7049) \quad (0.0744) \qquad r^2 = 0.8887 \qquad RSS_1 = 377.17 \qquad df = 11$$

الانحدار وفقًا للـ 13 مشاهدة الأخيرة

$$\hat{Y}_i = 28.0272 + 0.7941X_i$$

$$(30.6421) \quad (0.1319) \qquad r^2 = 0.7681 \qquad RSS_2 = 1536.8 \qquad df = 11$$

من هذه النتائج نحصل على

$$\lambda = \frac{RSS_2/df}{RSS_1/df} = \frac{1536.8/11}{377.17/11}$$

$$\lambda = 4.07$$

قيمة F الحرجة بدرجات حرية البسط 11 والمقام 11 عند مستوى معنوية 5% تساوي 2-82. بما أن قيمة F المقدرة (=λ) تزيد عن القيمة الحرجة، يمكن أن نستنتج أنه يوجد اختلاف في التباين الخاص بالخطأ. عمومًا، إذا ثبتنا مستوى المعنوية عند 1%، قد لا ترفض فرض ثبات التباين. (لماذا؟)، لاحظ أن قيمة P-valre لـ λ المشاهدة تساوي 0.014.

جدول (3.11) بيانات افتراضية عن نفقات الاستهلاك Y ($) والدخل X ($) لتوضيح اختبار Goldfeld - Quandt

Hypothetical Data on consumption Expenditine Y ($) and incene X ($) to illustrate the Goldfeld - Quandt test

		Data ranked by *X* values	
Y	*X*	*Y*	*X*
55	80	55	80
65	100	70	85
70	85	75	90
80	110	65	100
79	120	74	105
84	115	80	110
98	130	84	115
95	140	79	120

تابع - جدول (3.11) بيانات افتراضية عن نفقات الاستهلاك Y ($) والدخل X ($) لتوضيح اختبار Goldfeld - Quandt

		Data ranked by X values		
Y	X	Y	X	
90	125	90	125	
75	90	98	130	
74	105	95	140	
110	160	108	145	
113	150	113	150	
125	165	110	160	Middle 4 observations
108	145	125	165	
115	180	115	180	
140	225	130	185	
120	200	135	190	
145	240	120	200	
130	185	140	205	
152	220	144	210	
144	210	152	220	
175	245	140	225	
180	260	137	230	
135	190	145	240	
140	205	175	245	
178	265	189	250	
191	270	180	260	
137	230	178	265	
189	250	191	270	

اختبار Breusch-Pagan-Godfrey [21] :

نجاح اختبار Goldfeld-Quandt لا يعتمد فقط على قيمة c (عدد المشاهدات المركزية التي يتم حذفها) ولكن يعتمد أيضًا على التعريف السليم للمتغير X الذي يتم ترتيب المفردات وفقًا له. هذه القيود الخاصة بذلك الاختبار يمكن تجنبها إذا استخدمنا اختبار Breusch-Pagan-Godfrey (BPG) لشرح هذا الاختبار، دعنا نعتبر نموذج إنحدار يشمل على k متغير كالتالي:

$$Y_i = b_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \qquad (12.5.11)$$

افترض أن تباين الخطأ σ_i^2 معرف كالتالي:

$$\sigma_i^2 = f(\alpha_1 + \alpha_2 Z_{2i} + \dots + \alpha_m Z_{mi}) \qquad (13.5.11)$$

(21) T. Breusch and A. Pagan, "A Simple Test for Heteroscedasticity and Random Coefficient Variation," Econometrica, vol. 47, 1979, pp. 1287–1294.

انظر أيضًا في L. Godfrey, "Testing for Multiplicative Heteroscedasticity," Journal of Econometrics, vol. 8, 1978, pp. 277–236. Because of similarity, these tests are known as Breusch-Pagan-Godfrey tests of heteroscedasticity.

أي أن σ_i^2 يعتبر دالة ما في بعض المتغيرات غير العشوائية Z's، بعض الـ X's يمكن التعامل معها كالـ Z's. وبالأخص افترض أن:

$$\sigma_i^2 = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi} \qquad \textbf{(14.5.11)}$$

أي أن σ_i^2 هو دالة خطية في الـ Z's. إذا كان $\alpha_2 = \alpha_3 = ... = \alpha_m = 0$ فإن $\sigma_i^2 = \alpha_1$ وهي ثابت. بالتالي لاختبار ما إذا كانت σ_i^2 ثابتة التباين، يمكن أن نختبر أن $\alpha_2 = \alpha_3 = ... = \alpha_m = 0$. وهذه هي الفكرة الأساسية وراء اختبار Breusch-Pagon.

الاختبار الفعلي يتبع الخطوات التالية:

خطوة 1. قدر (12.5.11) باستخدام OLS واحصل على البواقي $\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n$.

خطوة 2. احصل على $\tilde{\sigma}^2 = \Sigma \hat{u}_i^2 / n$. تذكر أنه من (الفصل 4) ذلك التقدير هو تقدير الإمكان الأعظم الـ σ^2 (MLE). (لاحظ أن مقدر OLS هو $(\Sigma \hat{u}_i^2/(n-k)$).

خطوة 3. كون المتغير p_i كالتالي:

$$p_i = \hat{u}_i^2 / \tilde{\sigma}^2$$

والذي ببساطة يعبر عن مربعات البواقي مقسومة على $\tilde{\sigma}^2$.

خطوة 4. قم بعمل انحدار لـ p_i على الـ Z's

$$p_i = \alpha_1 + \alpha_2 Z_{2i} + \ldots + \alpha_m Z_{mi} + v_i \qquad \textbf{(15.5.11)}$$

حيث v_i يمثل حد البواقي في هذا الانحدار.

خطوة 5. احصل على ESS (مجموع المربعات المفسرة) من (15.5.11) وعرف التالي:

$$\Theta = \frac{1}{2}(\text{RSS})$$

بافتراض أن u_i يتبع التوزيع الطبيعي، من الممكن إثبات أنه إذا كان هناك ثبات في التباين ومع افتراض زيادة حجم العينة n، فإن

$$\Theta \underset{\text{asy}}{\sim} \chi^2_{m-1} \qquad \textbf{(17.5.11)}$$

أي أن، Θ تتبع توزيع كاي - التربيعي بدرجات حرية $(m-1)$. (لاحظ أن asy تعني تقاربيًا).

وبالتالي إذا وجد الباحث أن قيمة Θ المحسوبة $(= X^2)$ تزيد عن القيمة X_2 الحرجة عند مستوى المعنوية المحدد، فإنه يمكن رفض فرض ثبات التباين، وبخلاف ذلك لا يتم رفض الفرض.

وقد يتساءل الباحث لماذا اختار BPG القيمة $\frac{1}{2}$ ESS كإحصاء الاختبار؟ الإجابة معقدة نوعًا ما وموجودة في المراجع.[22]

مثال 5.11

اختبار BREUSCH - PAGAN - GODFREY (BPG)

كمثال دعنا نعيد استخدام بيانات (جدول 3.11) والتي تم استخدمها بعمل انحدار لـ Y على X، نحصل على التالي:

خطوة 1

$$\hat{Y}_i = 9.2903 + 0.6378X_i$$
$$\text{se} = (5.2314) \quad (0.0286) \qquad \text{RSS} = 2361.153 \qquad R^2 = 0.9466 \qquad (18.5.11)$$

خطوة 2

$$\tilde{\sigma}^2 = \sum \hat{u}_i^2/30 = 2361.153/30 = 78.7051$$

خطوة 3 . قسّم مربعات البواقي $\hat{u}_i$ التي حصلنا عليها من انحدار (18.5.11) على 78.7051 لتكوين المتغير p_i.

خطوة 4 . افترض أن p_i مرتبط خطيًا مع X_i ($=Z_i$) كما في (14.5.11) وبالتالي نحصل على الانحدار التالي:

$$\hat{p}_i = -0.7426 + 0.0101X_i$$
$$\text{se} = (0.7529) \quad (0.0041) \qquad \text{ESS} = 10.4280 \qquad R^2 = 0.18 \qquad (19.5.11)$$

خطوة 5

$$\Theta = \frac{1}{2}(\text{ESS}) = 5.2140 \qquad (20.5.11)$$

وفقًا لفروض اختبار BPG فإن Θ الموجودة في (20.5.11) تؤول تقاربيًا إلى توزيع كاي - التربيعي بدرجة حرية 1 . [لاحظ : يوجد متغير منحدر واحد فقط (19.5.11)] . والآن باستخدام جداول كاي التربيعية وعند درجة حرية 1 وباستخدام 5% مستوى معنوية غير أن قيمة كاي التربيعية الحرجة هي 3.8414 وقيمها عند 1% مستوى معنوية هي 6.6349 . وبالتالي فإن قيمة كاي التربيعية المحسوبة والتي تساوي 5.2140 لها معنوية إحصائية عند مستوى 5% ولكنها غير معنوية عند مستوي معنوية 1% . وبالتالي نصل إلى نفس الاستنتاج الذي توصل إليه اختبار Goldfeld - Quandt . ولكن دعنا نضع في الاعتبار وبشكل صارم، أن اختبار BPG هو اختبار تقاربي، أو يجب أن يستخدم مع أحجام العينات الكبيرة، اختبار العينة الحالية والمكونة من 30 مفردة قد لا تكون كافية لاعتبارها

(22) انظر Adrian C. Darnell, A Dictionary of Econometrics, Edward Elgar, Cheltenham, U.K., 1994, pp. 178–179.

عينة كبيرة الحجم. ويجب ملاحظة أيضًا أنه في العينات صغيرة الحجم يكون الاختبار متأثراً بشدة بفرض اعتيادية الأخطاء u_i. بالطبع من الممكن اختبار فرض الاعتيادية باستخدام الفروض التي تم استعراضها من قبل في (الفصل 5).[23]

الاختبار العام لـ White لاختلاف التباين: White's General Heteroscedasticity Test

بخلاف اختبار Goldfeld-Quandt والذي يتطلب إعادة ترتيب المشاهدات وفقًا للمتغير X، والذي يفترض أنه بسبب اختلاف التباين، وبخلاف أيضًا اختبار BPG والذي يتأثر بسرعة بفرض الاعتيادية، يأتي الاختبار العام لاختلاف التباين، والمقترح بواسطة White والذي لا يعتمد على فرض الاعتيادية ويسهل تطبيقه.[24]

ولشرح الفكرة الأساسية وراء هذا الاختبار، دعنا نعتبر نموذج الانحدار التالي والذي يشتمل على ثلاثة متغيرات (الحالة العامة والتي تتمثل في نموذج يشتمل على k متغير هي امتداد وتضييق مباشرة لهذه الحالة الخاصة).

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad \textbf{(21.5.11)}$$

خطوات اختبار White تتم كالتالي:

خطوة 1. وفقًا للبيانات نقوم بتقدير (21.5.11) ونحصل على البواقي $\hat{u}_i$.

خطوة 2. بعد ذلك نقوم بعمل الانحدار (المساعد) التالي:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}^2 + \alpha_5 X_{3i}^2 + \alpha_6 X_{2i} X_{3i} + v_i$$

(22.5.11)[25]

أي نقوم بعمل انحدار لمربعات البواقي من الانحدار الأصلي على متغيرات X الأصلية أو المنحدرة، ومربعاتها وحاصل ضربها. يمكن استخدام أس أعلى من 2 للمتغيرات المنحدرة. لاحظ أنه يوجد مقدار ثابت في الانحدار الأصلي. احصل على R^2 من هذا الانحدار (المساعد).

(23) لتوضيح ذلك انظر في R. Koenker, "A Note on Studentizing a Test for Heteroscedasticity," Journal of Econometrics, vol. 17, 1981, pp. 1180–1200.

(24) H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity," Econometrica, vol. 48, 1980, pp. 817–818.

(25) تتضمن هذه الطريقة فرضًا خاصًا بتباين الأخطاء σ_i^2 وهو أنه دالة في المتغيرات المنحدرة ومربعاتها وحاصل ضربها. إذا كانت كل التفاضلات الجزيئية في هذا الانحدار تساوى الصفر آنيًا فإن متغير الخطأ ثابت التباين وهذا الثابت يساوى α_1.

خطوة 3. تحت صحة الفرض العدمي والقائل بأنه لا يوجد اختلاف في التباين يمكن إثبات أنه حاصل ضرب حجم العينة (n) مع R^2 التي حصلنا عليها من الانحدار المساعد تؤول تقاربيًا إلى توزيع كاي - التربيعي بدرجات حرية تساوي عدد المتغيرات المنحدرة (مع استبعاد الجزء الثابت) في الانحدار المساعد أي أن:

$$n \cdot R^2 \underset{\text{asy}}{\sim} \chi^2_{\text{df}} \qquad (23.5.11)$$

حيث df معرفة كما سبق وحددناها. في مثالنا الحالي، يوجد 5 درجات حرية، حيث إنه يوجد 5 متغيرات منحدرة في الانحدار المساعد.

خطوة 4. إذا كانت قيمة كاي التربيعية التي حصلنا عليها من (23.5.11) تزيد عن قيمة كاى التربيعية عند مستوى المعنوية المحدد، فإننا نستنتج وجود اختلاف في التباين. وإذا لم تزد عن قيمة كاي التربيعية الحرجة فإنه لا يوجد اختلاف في التباين، أي انه في الانحدار المساعد (21.5.11) نجد أن:

$$\alpha_2 = \alpha_3 = \alpha_4 = \alpha_5 = \alpha_6 = 0$$

(انظر الهامش 25)

مثال 6.11

اختبار WHITE لاختلاف التباين : WHITE'S HETEROSCEDASTICITY TEST

من البيانات المقطعية الخاصة بـ 41 دولة. حصل Stephen Lewrs على نموذج الانحدار التالي:[26]

$$\ln Y_i = \beta_1 + \beta_2 \ln X_{2i} + \beta_3 \ln X_{3i} + u_i \qquad (24.5.11)$$

حيث Y = نسبة الضرائب التجارية (ضرائب الصادرات والواردات) إلى إجمالي الربح الحكومي، X_2 = نسبة مجموع الصادرات مع الواردات إلى GNP، و X_3 = GNP للفرد وذلك في صورة اللوغاريتم الطبيعي. افتراضه يعتمد على أن Y وX_2 مرتبطان طرديًا (أي زاد حجم التجارة، كلما زاد ربح الضرائب التجارية) و Y و X_3 مرتبطان عكسيًا (أي كلما زاد الدخل، فإن الحكومة ترى أنه من الأسهل تجميع ضرائب مباشرة - مثل ضرائب الدخل بدلاً من الاعتماد على الضرائب التجارية).

(26) Stephen R. Lewis, "Government Revenue from Foreign Trade," Manchester School of Economics and Social Studies, vol. 31, 1963, pp. 39–47.

النتائج التطبيقية دعمت هذه الافتراضات. وبالنظر إلى موضوعنا محل الدراسة، فإن النقطة المهمة هي هل يوجد اختلاف في التباين في هذه البيانات أم لا؟. بما أن البيانات تمثل بيانات مقطعية تحتوى العديد من الدول، فإنه مسبقًا يتصور الباحث وجود اختلاف في تباين الأخطاء.

بتطبيق اختبار White لاختلاف التباين على البواقي التي حصلنا عليها من انحدار (24.5.11) نحصل على النتائج التالية:[27]

$$\begin{aligned}\widehat{u_i^2} = &-5.8417 + 2.5629 \ln \text{Trade}_i + 0.6918 \ln \text{GNP}_i \\ &-0.4081(\ln \text{Trade}_i)^2 - 0.0491(\ln \text{GNP}_i)^2 \\ &+0.0015(\ln \text{Trade}_i)(\ln \text{GNP}_i) \qquad R^2 = 0.1148\end{aligned} \tag{25.5.11}$$

لاحظ أن: الأخطاء القياسية ليست معطاة، حيث إنها غير مرتبطة بهدف دراستنا الحالية.

الآن نجد أن $n{\cdot}R^2 = 41\ (0.1148) = 4.7068$، وهذه القيمة تؤول تقاربيًا إلى توزيع كاي التربيعي بدرجات حرية تساوي 5 (لماذا؟) قيمة كاى التربيعية الحرجة عند مستوى معنوية 5% وبدرجات حرية 5 هي 11.0705، عند 10% تساوي 9.2363 والـ 25% قيمة حرجة هي 6.62568. وبالتالي باستخدام كل هذه القيم، يمكن للباحث أن يستنتج أنه وفقًا لاختيار White لا يوجد اختلاف في التباين.

هناك تعليق مهم وخاص باختبار White، إذا كان النموذج يشتمل على العديد من المتغيرات المنحدرة، فإن استخدام كل هذه المتغيرات مع مربعاتها (أو أس أعلى) بالإضافة إلى حواصل ضربها قد يستهلك درجات حرية أكبر. وبالتالي يجب الحرص من هذه النقطة عند استخدام هذا الاختبار.[28]

في الحالات التي يكون فيها إحصاء اختبار White المعطى في (25.5.11) معنويًا احصائيًا، قد لا يكون اختلاف التباين هو بالضرورة السبب في ذلك، ولكن الخطأ في التوصيف قد يلعب دورًا في ذلك، وسنناقش ذلك بالتفصيل في (الفصل 13) (تذكر النقطة 5 في الفقرة 1.11). بمعنى آخر، اختبار White يمكن أن يكون اختبارًا (خاصًا) باختلاف التباين أو خطأ التوصيف أو كليهما معًا. تم مناقشة أيضًا أنه إذا كانت البيانات ليست بيانات مقطعية، فإن اختبار White هو اختبار خاص فقط

---

(27) هذه النتائج، مع تغيير بعض الرموز، مقدمة في
William F. Lott and Subhash C. Ray, Applied Econometrics: Problems with Data Sets, Instructor's Manual, Chap. 22, pp. 137–140.

(28) أحيانًا من الممكن تعديل هذا الاختبار لاستخدام درجات حرية منطقية انظر تمرين 18.11 .

وبشكل خاص لاختلاف التباين. ولكن إذا كانت البيانات محل الدراسة بيانات مقطعية فإنه يعتبر اختبار لكل من اختلاف التباين تحيز التوصيف.[29]

اختبارات أخرى لاختلاف التباين: Other Tests of Heteroscedasticity

هناك العديد من الاختبارات الأخرى الخاصة باختلاف التباين، وكل منهما يعتمد على افتراضات معينة. يمكن للقارئ المهتم بذلك أن يرجع إلى قائمة المراجع.[30] سنذكر أحد هذه الاختبارات لسهولته. هذا الاختبار هو اختبار (KB) Koenker-Bassett. مثل اختبارات Park-Breusch-Pagan-Godfrey واختبار White للاختلاف التباين، فإن اختبار KB يعتمد على مربعات البواقي، $\hat{u}_i^2$، ولكن بدلاً من عمل الانحدار على متغير أو أكثر من المتغيرات المنحدرة، فإن انحدار مربعات البواقي يكون على مربعات القيم المقدرة للمتغير المنحدر عليه. وبالتالي دعنا نفترض أن النموذج الأصلي كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \qquad \textbf{(26.5.11)}$$

فإننا نقدر هذا النموذج، ونحصل على $\hat{u}_i$ من النموذج، ثم نقدر الانحدار التالي:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 (\hat{Y})^2 + v_i \qquad \textbf{(27.5.11)}$$

حيث $\hat{Y}_i$ هي القيم المقدرة من النموذج (26.5.11). الفرض العدمي الخاص بأن $\alpha_2 = 0$ إذا لم نرفضه، يمكن أن نستنتج أنه لا يوجد اختلاف في التباين. الفرض العدمي يمكن اختياره باختبار t المعتاد أو اختبار F. (لاحظ أن $F_{1,k} = t_k^2$)

إذا كان النموذج (26.5.11) هو نموذج لوغاريتمي مضاعف، فإن مربعات البواقي ستنحدر على $(\log \hat{Y}_i)^2$. إحدى مميزات اختبار KB أنه يمكن تطبيقه حتى إذا كانت الأخطاء من النموذج الأصلي (26.5.11) لا تتبع التوزيع الطبيعي. إذا طبقنا اختبار KB على مثال 1.11، سنجد أن معامل الميل في انحدار مربعات البواقي الذي

---

(29) انظر Richard Darris, Using Cointegration Analysis in Econometrics Modelling, Prentice Hall & Harvester Wheastsheaf, U.K., 1995, p. 68.

(30) انظر M. J. Harrison and B. P. McCabe, "A Test for Heteroscedasticity Based on Ordinary Least Squares Residuals," Joumal of the American Statistical Association, vol 74, 1979, pp. 494–499; J. Szroeter, "A Class of Parametric Tests for Heteroscedasticity in Linear Econometric Models, "Econometrica, vol 46, 1978, pp. 1311–1327; M. A. Evans and M. L. King, "A Further Class of Tests for Heteroscedasticity," Journal of Econometrics, vol. 37; 1988 pp. 265–276; R. Koenker and G. Bassett, "Robust Tests for Heteroscedasticity Based on Regression Quantiles," Econometrica, vol. 50, 1982, pp.43–61.

نحصل عليه من (3.5.11) على قيم المقدرة من (3.5.11) لا يختلف إحصائيًا عن الصفر، أي أنه يدعم اختبار Park. هذه النتيجة يجب ألا تكون نتيجة مفاجئة حيث إننا لدينا متغيرًا منحدرًا واحدًا فقط. ولكن اختبار KB يمكن تطبيقه إذا كان لدينا متغير منحدر واحد أو أكثر.

6.11 المقاييس العلاجية: REMEDIAL MEASURES

كما سبق ورأينا، فإن اختلاف التباين لا يؤثر على خواص عدم التحيز أو الاتساق الخاصة بمقدرات الـ OLS، ولكن يؤثر فقط على خاصية الكفاءة ولكن ليس على المستوى التقاربي (أي مع أحجام العينات الكبيرة). عدم تميز المقدرات بالكفاءة يجعل خطوات اختبارات الفروض التقليدية ملتبسة ومشكوكاً في قيمها. وبالتالي فإن المقاييس العلاجية قد يكون لها دور في ذلك. هناك طريقتان للعلاج: طريقة يمكن استخدامها عندما تكون σ_i^2 معلومة، وطريقة عندما تكون σ_i^2 غير معلومة.

عندما تكون σ_i^2 معلومة: طريقة المربعات الصغرى المرجحة:

When σ_i^2 is Known: The Method of Weighted Least Squares

كما سبق ورأينا في الفقرة 3.11، إذا كانت σ_i^2 معلومة، فإن أكثر الطرق مباشرة لتصحيح اختلاف التباين هي المربعات الصغرى المرجحة، وبالتالي نحصل على مقدرات تكون BLUE.

مثال 7.11

شرح طريقة المربعات الصغرى المرجحة:

ILLUSTRATION OF THE METHOD OF WEIGHTED LEAST SQUARES

لشرح هذه الطريقة، دعنا نفترض أننا نريد دراسة العلاقة بين التعويضات وحجم العمالة وفقًا للبيانات الموجودة في جدول 1.11. للتبسيط سنقيس حجم العمالة بـ 1 (1-4 عامل)، 2 (2-9 عامل)، . . . ، 9 (1000 - 2499 عامل)، ومن الممكن أيضًا أن نقيس الفترة بالقيمة الوسطى لفترة العمالة المعطاة في الجدول. دعنا نعتبر Y هي متوسط التعويضات بالنسبة للعامل ($) و X هي حجم العمالة سنقوم بعمل الانحدار التالي [انظر المعادلة (6.3.11)]:

$$Y_i/\sigma_i = \hat{\beta}_1^*(1/\sigma_i) + \hat{\beta}_2^*(X_i/\sigma_i) + (\hat{u}_i/\sigma_i) \quad (1.6.11)$$

حيث σ_i هي الانحراف المعياري للأجور كما هي مسجلة في جدول (1.11). البيانات الخام المطلوبة لإجراء هذا الانحدار معطاة في جدول (4.11).

قبل الانتقال إلى نتائج الانحدار، لاحظ أن (1.6.11) لا يشتمل على جزء ثابت مقطوع من المحور الصادي (لماذا؟). وبالتالي، فإننا سنستخدم الانحدار الذي يمر بنقطة الأصل لتقدير β_1^* وβ_2^* وهذا الموضوع تمت مناقشته من قبل في (الفصل 6). ولكن معظم برامج الحاسب الآلي الآن لديها إمكانية لعدم تفعيل الجزء المقطوع من المحور الصادي (انظر Minitab أو Eviews) لاحظ أيضًا خاصية مهمة أخرى متعلقة بـ (1.6.11)، فلدينا متغيران مفسران $(\frac{1}{\sigma_i})$ و $(\frac{X_i}{\sigma_i})$، في حين أننا إذا كنا سنستخدم OLS لانحدار التعويضات على حجم العمالة، فإن هذا الانحدار سيشتمل على متغير مفسر واحد فقط وهو X_i (لماذا؟).

نتائج الانحدار الخاصة بـ WLS كالتالي:

$$\widehat{(Y_i/\sigma_i)} = 3406.639(1/\sigma_i) + 154.153(X_i/\sigma_i)$$
$$(80.983) \qquad (16.959) \qquad \textbf{(2.6.11)}$$
$$t = (42.066) \qquad (9.090)$$
$$R^2 = 0.9993^{(31)}$$

للمقارنة، دعنا نستعرض نتائج انحدار OLS العادية أو غير المرجحة:

$$\hat{Y}_i = 3417.833 + 148.767\,X_1 \qquad \textbf{(3.6.11)}$$
$$(81.136) \qquad (14.418)$$
$$t = (42.125) \qquad (10.318) \qquad R^2 = 0.9383$$

في تمرين 7.11 سيطلب من القارئ أن يجري مقارنة بين هذين الانحدارين.

جدول (4.11) شرح انحدار المربعات الصغرى المرجحة **Illustration of Weighted Least - Squares Regression**

Compensation, Y	Employment size, X	σ_i	Y_i/σ_i	X_i/σ_i
3396	1	743.7	4.5664	0.0013
3787	2	851.4	4.4480	0.0023
4013	3	727.8	5.5139	0.0041
4104	4	805.06	5.0978	0.0050
4146	5	929.9	4.4585	0.0054
4241	6	1080.6	3.9247	0.0055
4387	7	1243.2	3.5288	0.0056
4538	8	1307.7	3.4702	0.0061
4843	9	1112.5	4.3532	0.0081

لاحظ أن: في انحدار (2.6.11)، المتغير التابع هو (Y_i/σ_i) والمتغيرات المستقلة هي $(\frac{1}{\sigma_i})$، $(\frac{X_i}{\sigma_i})$.

المصدر: بيانات خاصة بـ Y و σ_i (الانحراف المعياري للتعويضات) من جدول (3.11)، حجم العمالة 1=1–4 عمال، 2=5–9 عمال وهكذا، البيانات الأخيرة أيضًا تم الحصول عليها من جدول (1.11).

(31) كما لاحظنا في الهامش 3 في الفصل 6، قيمة R^2 الخاصة بالانحدار المار بنقطة الأصل لا يمكن مقارنته مباشرة مع R^2 التي نحصل عليها من نموذج به جزء مقطوع من المحور الصادي. قيمة R^2 المسجلة والمساوية تصحيح R^2 لتأخذ في اعتبارها غياب الجزء المقطوع من المحور الصادي، انظر أيضا App. 1A6 الفقرة 1A6).

عندما تكون σ_i^2 غير معلومة : When σ_i^2 is not Known

كما سبق وذكرنا، عندما تكون σ_i^2 معلومة، فإننا نستخدم طريقة WLS للحصول على المقدرات الـ BLUE . وحيث إنه نادراً ما تكون σ_i^2 معلومة، فهل هناك طريقة يمكن الحصول منها على مقدر متسق (بالمعنى الإحصائي) لتباين وتغاير مقدرات الـ OLS إذا كان هناك اختلاف في التباين؟ الإجابة هي نعم .

التباينات والأخطاء القياسية المتسقة لـ White – في حالة اختلاف التباين :
White's Heteroscedasticity-Consistent Variances and Standard Errors

وضح White أن هذا المقدر له خواص تقاربية جيدة (أي مع أحجام العينات الكبيرة) حيث يمكن عمل استدلال إحصائي حول القيم الحقيقية للمعالم .[32]

لن نستعرض هنا التفاصيل الرياضية، حيث إنه خارج نطاق الكتاب الحالي، وعموماً الخطوط العريضة للطريقة تم توضيحها في ملحق 4.A11 . الآن يوجد العديد من حزم الحاسب الآلي التي تستطيع القيام بحساب مقدر White للتبين والأخطاء القياسية المصحح لاختلاف التباين في إطار حساب التباينات والأخطاء القياسية المعدلة بـ OLS التقليدية .[33]

وتعرف أيضاً الأخطاء القياسية المعدلة لـ white في حالة وجود اختلاف في التباين باسم الأخطاء القياسية الثابتة (اللامعلمية) .

مثال 8.11

شرح طريقة ILLUSTRATION OF WHITE'S PROCEDURE : WHITE

كمثال، دعنا نستعرض نتائج Greene كالتالي :[34]

$$\hat{Y}_i = 832.91 - 1834.2\ (\text{income}) + 1587.04\ (\text{income})^2$$

$$\begin{aligned} \text{OLS se} &= (327.3) \quad (829.0) \quad (519.1) \qquad (4.6.11)\\ t &= (2.54) \quad (2.21) \quad (3.06)\\ \text{White se} &= (460.9) \quad (1243.0) \quad (830.0)\\ t &= (1.81) \quad (-1.48) \quad (1.91) \end{aligned}$$

حيث Y = الإنفاق على التعليم الحكومي بالنسبة للفرد وفقاً للولاية في 1979 و Incone = دخل الفرد وفقاً للولاية في 1979 . البيانات خاصة بـ 50 ولاية بالإضافة إلى ولاية واشنطن العاصمة .

(32) انظر H. White, op. cit.

(33) فنياً تعرف هذه المقدرات بمصفوفات مقدرات التغاير المتسقة في حالة وجود اختلاف في التباين .

(34) William H. Greene, Econometric Analysis, 2d ed., Macmillan, New York, 1993, p. 385

كما نرى من النتائج السابقة، الأخطاء القياسية المعدلة لـ white في حالة وجود اختلاف في التباين، أكبر من الأخطاء القياسية الخاصة بالـ OLS، وبالتالي فإن قيم t المقدرة الخاصة بها ستكون أصغر كثيراً من نظيرها الخاص بالـ OLS ووفقاً لهذه الملاحظة الأخيرة، سيكون كل من المتغيرين المنحدرين معنويين إحصائياً عند مستوى معنوية 5%، أما وفقاً لمقدرات White فالمتغيران المنحدران غير معنويين.

بوجه عام، يجب ملاحظة أن الأخطاء القياسية المعدلة لـ White في حالة وجود اختلاف في التباين، قد تكون أكبر أو أصغر من الأخطاء القياسية غير المعدلة.

وحيث إن مقدرات White المتسقة والتي يمكن استخدامها في حالة اختلاف التباين متوافرة الآن في معظم الحزم الإحصائية الخاصة بالانحدار، فإن من المفيد للقارئ أن يستخدمها. وكما قال كل من Wallace و Silver note:

بوجه عام، يعتبر استخدام مقدرات White (المتاحة في معظم برامج الانحدار) أمراً مفيداً، حيث يمكن مقارنة النتائج مع نتائج OLS التقليدية للتأكد مما إذا كان اختلاف التباين يمثل مشكلة حقيقية في البيانات محل الدراسة أم لا.[35]

فروض ظاهرية خاصة بنمط اختلاف التباين :
Plausible Assumptions about heteroscedasticity pattern

بالإضافة إلى أن طريقة White تعتبر طريقة خاصة لأحجام العينات الكبيرة فقط، إلا أن أحد عيوبها أيضاً هو أن المقدرات التي نحصل عليها وفقاً لهذه الطريقة قد تكون مقدرات غير كفء، حيث إنه تم الحصول عليها من بيانات محولة لتعكس نوعاً معيناً من اختلاف التباين. لشرح ذلك، دعنا نسترجع نموذج الانحدار الذي يشتمل على متغيرين اثنين فقط:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

دعنا الآن نفترض عدداً من الفروض الخاصة بنمط اختلاف التباين:

الفرض الأول: تباين الخطأ يتناسب مع X_i^2:
$E(u_i^2) = \sigma^2 X_i^2$ **(5.6.11)**[36]

(35) T. Dudley Wallace and J. Lew Silver, Econometrics: An Introduction, Addison-Wesley, Reading, Mass., 1988, p.265.

(36) تذكر أننا تعرضنا من قبل لهذا الفرض عندما ناقشنا اختبار Coldfeld-Quandt .

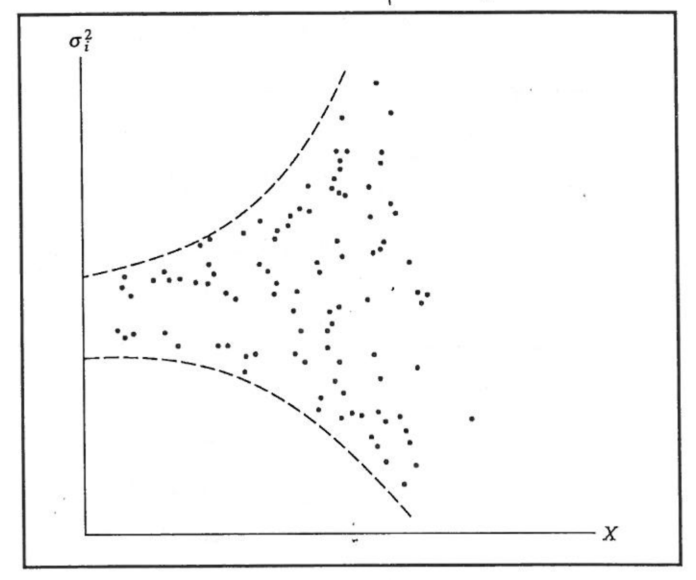
وفقًا للطرق البيانية أو طرق Park و Glejser فإن تباين u_i يتناسب مع مربع المتغير المفسر X [انظر شكل (10.11)] . ممكن أن يتم تحـويل النمـوذج الأسـاسي كـالتـالي : (نقسم النموذج الأساسي على X_i):

$$\frac{Y_i}{X_i} = \frac{\beta_1}{X_i} + \beta_2 + \frac{u_i}{X_i} = \beta_1 \frac{1}{X_i} + \beta_2 + v_i \qquad (6.6.11)$$

حيث v_i هو مقدار الخطأ المحول ويساوي u_i/X_i . الآن من السهل إثبات التالي :

$$E\left(v_i^2\right) = E\left(\frac{u_i}{X_i}\right)^2 = \frac{1}{X_i^2} E\left(u_i^2\right)$$

$= \sigma^2$ باستخدام (5.6.11)



شكل (10.11) تباين الخطأ يتناسب مع X^2 : **Error Variance Proportional to X^2**

وبالتـالي ، فإن تبـاين v_i يعتبر الآن متـجانسًا وثابتًا ، ويمكن للفـرد أن يطبق OLS على المعادلة المحولة (6.6.11) بعمل انحدار لـ Y_i/X_i على $1/X_i$.

لاحظ أنه في الانحدار المحول الجزء المقطوع من المحور الصادي β_2 هو معامل الميل في المعـادلـة الأصليـة ، ومـعـامـل الميل β_1 هو الجـزء المقطوع من المحـور الـصـادي في النمـوذج الأصلي ، وبالتـالي حتى نعـود مـرة أخرى إلى النمـوذج الأصلي ، يجب أن

نضرب المعادلة (6.6.11) المقدرة في X_i تمرين 20.11 يعتبر تطبيقًا لهذه التحويلة.

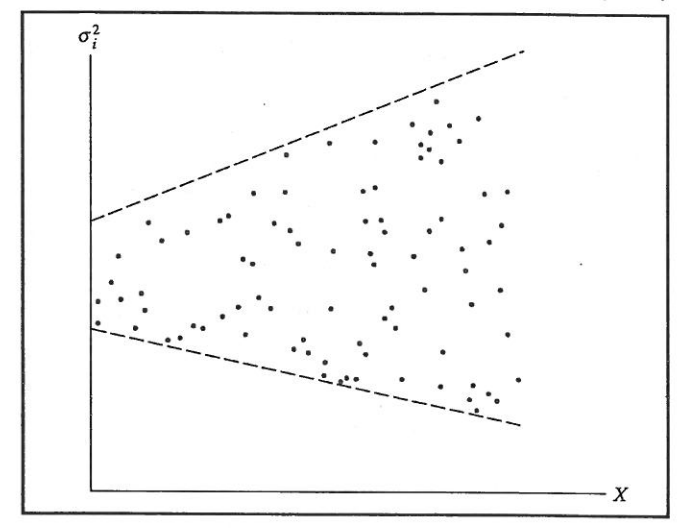
الفرض الثاني : تباين الخطأ يتناسب مع X_i. تحويلة الجذر التربيعي

$$E(u_i^2) = \sigma^2 X_i \tag{7.6.11}$$

هنا يعتبر تباين u_i، بدلاً من أن يكون متناسبًا مع مربع X_i، يعتبر متناسبًا مع X_i بنفسها. وبالتالي فإن النموذج الأصلي يتم تحويله كالتالي [انظر شكل (11.11)]

$$\begin{aligned} \frac{Y_i}{\sqrt{X_i}} &= \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}} \\ &= \beta_1\frac{1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + v_i \end{aligned} \tag{8.6.11}$$

حيث $v_i = u_i/\sqrt{X_i}$ و $X_i > 0$.



شكل (11.11) تباين الخطأ يتناسب مع X : Error Variance proportional to X

بناء على الفرض الثاني، من الممكن إثبات أن $E(v_i^2) = \sigma^2$ وهذا يعتبر ثباتًا للتباين، وبالتالي يمكن للباحث الآن أن يطبق OLS على (8.6.11) ويقوم بعمل انحدار لـ $Y_i/\sqrt{X_i}$ على $1/\sqrt{X_i}$ و $\sqrt{X_i}$.

لاحظ خاصية مهمة للنموذج المحول: لا يوجد فيه جزء ثابت مقطوع من المحور الصادي، وبالتالي لابد من عمل انحدار لنموذج يمر بنقطة الأصل لتقدير β_1 و β_2. ووفقًا لـ (8.6.11) نرجع مرة أخرى إلى النموذج الأصلي بعد ضرب (8.6.11) بـ $\sqrt{X_i}$.

الفرض الثالث: تباين الخطأ يتناسب مع مربع القيمة المتوقعة للـ Y.

$$E(u_i^2) = \sigma^2[E(Y_i)]^2 \qquad \textbf{(9.6.11)}$$

المعادلة (9.6.11) تفترض أن تباين u_i يتناسب مع مربع القيمة المتوقعة للـ Y [انظر الشكل (e8.11)]. الآن

$$E(Y_i) = \beta_1 + \beta_2 X_i$$

وبالتالي دخول النموذج الأصلي كالتالي:

$$\frac{Y_i}{E(Y_i)} = \frac{\beta_1}{E(Y_i)} + \beta_2 \frac{X_i}{E(Y_i)} + \frac{u_i}{E(Y_i)}$$
$$= \beta_1\left(\frac{1}{E(Y_i)}\right) + \beta_2 \frac{X_i}{E(Y_i)} + v_i \qquad \textbf{(10.6.11)}$$

حيث $v_i = u_i/E(Y_i)$، نلاحظ أن $E(v_i^2) = \sigma^2$، أي أن الخطأ v_i ثابت التباين وبالتالي فإن انحدار (10.6.11) مستوفي فرض ثبات التباين الخاص بنموذج الانحدار التقليدي.

التحويلة (10.6.11) غير ممكنة، حيث إن $E(Y_i)$ تعتمد على β_1 و β_2 وكلاهما غير معلوم. بالطبع نحن نعرف أن $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$ وهو مقدر للـ $E(Y_i)$ وبالتالي يمكن أن نقوم بالخطوتين التاليتين: أولاً، نقوم بعمل انحدار OLS التقليدي، متجاهلين مشكلة اختلاف التباين، ونحصل على $\hat{Y}_i$. وبالتالي باستخدام المقدرة، نحول نموذجنا كالتالي:

$$\frac{Y_i}{\hat{Y}_i} = \beta_1\left(\frac{1}{\hat{Y}_i}\right) + \beta_2\left(\frac{X_i}{\hat{Y}_i}\right) + v_i \qquad \textbf{(11.6.11)}$$

حيث $v_i = \frac{u_i}{\hat{Y}_i}$. في الخطوة الثانية، نقوم بعمل انحدار (11.6.11). رغم أن $\hat{Y}_i$ لا تتساوى بالضبط مع $E(Y_i)$، إلا أنها مقدرات متسقة، أي أنها مع زيادة حجم العينة يؤول تقاربيًا إلى قيمة $E(Y_i)$ الحقيقية.

وبالتالي فإن التحويلة (11.6.11) سيكون تطبيقها مرضيًا في الواقع إذا كان حجم العينة كبيرًا بشكل كاف.

الفرض الرابع: تحويلة اللوغاريتم كالتالي:

$$\ln Y_i = \beta_1 + \beta_2 \ln X_i + u_i \qquad \textbf{(12.6.11)}$$

هذه التحويلة تقلل اختلاف التباين عندما يقارن مع الانحدار $Y_i = \beta_1 + \beta_2 X_i + u_i$.

هذه النتيجة تحدث لأن التحويلة اللوغاريتمية تضغط المقياس الذي نقيس به المتغيرات. وبالتالي يقل الفرق بين أي قيمتين ليصبح فرقًا أقل. فمثلاً القيمة 80 تساوي 10 مرات القيمة 8 ولكن In 80(=4.3280) يعتبر ضعف In 8 (=2.0794).

وميزة أخرى للتحويلة اللوغاريتمية، أن معامل الميل β_2 يقيس مرونة Y بالنسبة للـ X، أي النسبة التي يتغير بها Y وفقًا لنسبة التغير في X.

فمثلاً، إذا كانت Y هي الاستهلاك، و X الدخل، فإن β_2 الموجود في المعادلة (12.6.11) سيقيس مرونة الدخل. كما أنه في النموذج الأصلي فإن β_2 تقيس فقط معدل التغير في متوسط الاستهلاك بالنسبة للتغير بوحدة واحدة في الدخل. وهذا يعتبر أحد أسباب كثرة استخدم النماذج اللوغاريتمية في الاقتصاد القياسي التطبيقي [للتعرف على بعض المشاكل المرتبطة بالتحويلة اللوغاريتمية، انظر تمرين (4.11)].

لتلخيص مناقشتنا حول المقاييس العلاجية، أوضحنا أن كل التحويلات السابقة تعتمد على الموضوع محل التطبيق، وتدور حول طبيعة σ_i^2. ومن المناقشة السابقة أوضحنا أنها تعتمد على طبيعة المشكلة محل الدراسة، ودرجة الصعوبة الخاصة بمشكلة اختلاف التباين. **هناك بعض المشاكل الإضافية والمرتبطة بالتحويلات والتي وجدنا ضرورة وضعها في الاعتبار التالي:**

1 – عندما يشتمل النموذج على أكثر من متغيرين اثنين، قد لا نستطيع تحديد أي من متغيرات الـ X الذي يجب استخدامه لتحويل البيانات.[37]

2 – التحويلة اللوغاريتمية – كما سبق وناقشناها في الفرض 4 – لا يمكن تطبيقها إذا كانت بعض قيم الـ Y والـ X تساوي الصفر أو سالبة.[38]

3 – هناك أيضًا مشكلة الارتباط الزائف. هذا وفقًا لـ Karl Pearion يتمثل في الوضع الذي يظهر الارتباط بين نسب المتغيرات حتى لو كانت المتغيرات الأصلية غير

---

(37) عمومًا بشكل عملي يمكن للباحث أن يرسم $\hat{u}_i^2$ مع كل متغير ويحدد أيًا من متغيرات الـ X يمكن استخدامه لتحويل البيانات. (انظر شكل 9.11)

(38) أحيانًا يمكن استخدام $\ln (Y_i + k)$ أو $\ln (X_i + k)$، حيث k رقم موجب يختار بحيث يجعل كل قيم الـ Y والـ X موجبة.

مرتبطة أو عشوائية.[39] وبالتالي في النموذج $Y_i = \beta_1 + \beta_2 X_i + u_i$، Y و X قد لا يكونوا مرتبطين ولكن النموذج المحول $\frac{Y_i}{X_i} = \beta_1(\frac{1}{X_i}) + \beta_2$، $\frac{Y_i}{X_i}$ و $\frac{1}{X_i}$ سنجد أنهما غالبًا ما يكونان مرتبطين.

4 - إذا كانت σ_i^2 غير معلومة مباشرًا، وتقدر بواحدة أو أكثر من التحويلات السابق مناقشتها، فإن طريقة اختبارات الـ t والـ F وما إلى غير ذلك يجب أن تستخدم فقط في حالة أحجام العينات الكبيرة. وبالتالي يجب على الباحث أن يكون شديد الحرص عند تفسيره للنتائج المعتمدة على التحويلات المختلفة في العينات المحدودة أو صغيرة الحجم.[40]

7.11 أمثلة استنتاجية : CONCLUDING EXAMPLES

بعد مناقشتنا الحالية لاختلاف التباين، يمكن تقديم استنتاجاتنا في الأمثلة التالية، والتي تشرح النقاط المهمة التي تم استعراضها في هذا الفصل.

مثال 9.11

وفيات الأطفال (مرة أخرى) : CHILD MORTALITY REVISITED

دعنا نرجع مرة أخرى للمثال الخاص بوفيات الأطفال، والذي سبق واستخدمناه مرات عديدة من قبل. وفقًا للبيانات الخاصة بـ 64 دولة، حصلنا على نتائج الانحدار الموضحة في المعادلة (1.2.8). بما أن هذه البيانات تعتبر بيانات مقطعية، تشتمل على دول متنوعة ومختلفة وفقًا لمعدلات الوفيات، قد يكون محتمل بشكل كبير وجود اختلاف في التباين. وللتأكد من ذلك، دعنا نستعرض أولاً البواقي التي حصلنا عليها من المعادلة (1.2.8). هذه البواقي مرسومة في الشكل (2.11). من هذا الشكل يتضح أن البواقي لا يظهر فيها أي نمط يجعلنا نعتقد بأن هناك اختلافًا في التباين. فلاشيء يمكن استنتاجه من الشكل البياني. ولذلك دعنا نطبق اختبارات Park، Glejser و White لنرى ما إذا كان هناك أي دليل على وجود اختلاف في التباين.

---

(39) على سبيل المثال، إذا كانت X_1، X_2 و X_3 غير مرتبطة ثنائيًا أي أن $r_{12} = r_{13} = r_{23} = 0$ وعندما نجد أن (قيمته) نسب X_1/X_3 و X_2/X_3 مرتبطة، يكون هناك ما يسمى «بالارتباط الزائف» بوجه عام، فالارتباط يوصف بالزائف إذا وجد نتيجة لطريقة التعامل مع البيانات ولكن ليس موجودا في الشكل الأصلي لها. M. G. Kendall and W. R. Buckland, A Dictionary of Statistical Terms, Hafner Publishing, New York, 1972, p. 143.

(40) لمزيد من التفاصيل، انظر George G. Judge et al., op. cit., sec. 14.4, pp. 415–420

اختبار Park :

بما أن هناك متغيرين منحدرين اثنين، GNP و ELR، من الممكن أن نقوم بعمل انحدار للبواقي المربعة من انحدار (1.2.8) على أي من هذين المتغيرين، أو ممكن أن نقوم بعمل انحدار لهما على قيم CM المقدرة ($=\widehat{CM}$) من انحدار (1.2.8) باستخدام الأخير، وبذلك نحصل على النتائج التالية :

$$\widehat{\hat{u}_i^2} = 854.4006 + 5.8016\,\widehat{CM}_i \qquad (1.7.11)$$
$$t = \quad (1.2010) \quad (1.2428) \qquad r^2 = 0.024$$

لاحظ أن : $\hat{u}_i$ هي بواقي النموذج (1.2.8) و $\widehat{CM}$ هي القيم المقدرة للـ CM من نموذج (1.2.8)

يوضح هذا الانحدار، عدم وجود علاقة منتظمة بين مربعات البواقي والقيم المقدرة CM (لماذا؟)، مما يعني أن فرض ثبات التباين يمكن قبوله. ونلاحظ أيضًا أنه إذا قمنا بعمل انحدار للوغاريتم مربعات البواقي على لوغاريتم $\widehat{CM}$ فإن استنتاجنا لن يتغير.

اختبار Glejser :

القيم المطلقة للبواقي التي حصلنا عليها من (1.2.8)، عندما تم عمل انحدار لها على القيم المقدرة للـ CM من نفس الانحدار تجعلنا نحصل على النتائج التالية :

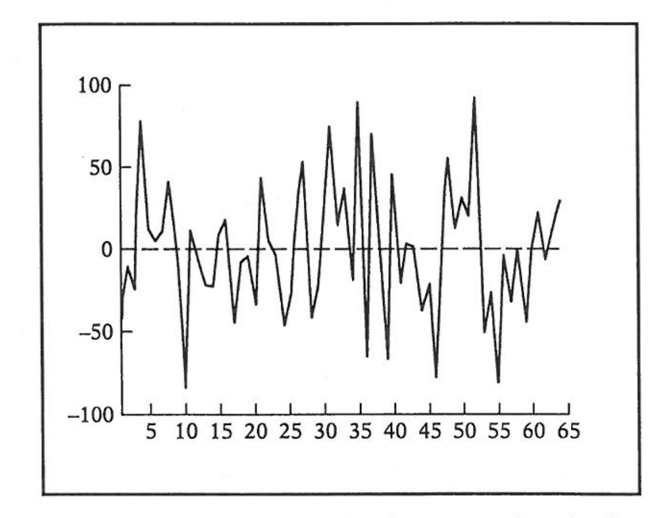
$$\widehat{|\hat{u}_i|} = 22.3127 + 0.0646\,\widehat{CM}_i \qquad (2.7.11)$$
$$t = \quad (2.8086) \quad (1.2622) \qquad r^2 = 0.0250$$

مرة أخرى، لا توجد علاقة منتظمة بين القيم المطلقة للبواقي والقيم المقدرة CM، حيث إن قيم t لمعامل الميل ليست لها معنوية إحصائية.

اختبار White :

عند تطبيق اختبار white لاختلاف التباين باستخدام حدود الضرب التبادلي، لم نجد أي دليل على وجود اختلاف في التباين وقمنا أيضًا بإعادة تقدير (1.2.8) للحصول على الأخطاء القياسية المتسقة لـ White في حالة اختلاف التباين وقيم t المرتبطة بها ولكن النتائج كانت مقاربة جدًا للموجودة في المعادلة (1.2.8)، وهذه النتيجة متوقعة وفقًا للنتائج السابقة التي حصلنا عليها من اختبارات اختلاف التباين السابقة.

في النهاية، يبدو بشكل واضح أن انحدار وفيات الأطفال (1.2.8) لا يعاني من مشكلة اختلاف التباين.



شكل (12.11) بواقي انحدار (1.2.8) Residuals from regression (8.2.1)

مثال 10.11

النفقات والمبيعات والأرباح لـ R&D خاصة بـ 18 مجموعة صناعية في الولايات المتحدة ، 1988

R&D EXPENDITURE, SALES, AND PROFITS IN 18 INDUSTRY GROUPINGS IN THE UNITED STATES, 1988

جدول (5.11) يعطي بيانات البحث والتطوير (R&D) للنفقات والمبيعات والأرباح الخاصة بـ 18 مجموعة صناعية في الولايات المتحدة . كل الأشكال البيانية مقاسة بالمليون دولار . بما أن البيانات المقطعية الموجودة في هذا الجدول توضح اختلافًا في التباين ، فإن انحدار R&D على المبيعات (أو الأرباح) غالبًا ما سيكون فيه مشكلة اختلاف التباين .

نتائج الانحدار كالتالي :

$$\widehat{R\,\&\,D}_i = 192.9931 + 0.0319\ \text{Sales}_i$$
$$\text{se} = (533.9317) \quad (0.0083) \qquad (3.7.11)$$
$$t = (0.3614) \quad (3.8433) \qquad r^2 = 0.4783$$

ونجد أن هناك علامة طردية معنوية بين R&D والمبيعات ، وتلك النتيجة تعتبر نتيجة متوقعة .

إذا أردنا معرفة ما إذا كان انحدار (3.7.11) يعاني من اختلاف التباين أم لا . يمكن أن نحصل على البواقي ، $\hat{u}_i$ ، ومربعات البواقي ، $\hat{u}_i^2$ ، من الانحدار السابق ونقوم برسمها مع المبيعات كما هو موضح في الشكل (3.11) . يبدو من الشكل البياني أن هناك نمطًا

محدداً بين البواقي ومربعات البواقي والمبيعات، مما يقترح وجود اختلاف في التباين.
لإجراء اختبار رسمي لذلك دعنا نستخدم اختبارات Park ، Gbejser و White والتي سيعطينا النتائج التالية:

اختبار Park :

$$\left|\widehat{\hat{u}_i^2}\right| = 974{,}469.1 + 86.2321\ \text{Sales}_i$$
$$\text{se} = (4{,}802{,}343) \quad (40.3625) \quad r^2 = 0.2219 \qquad \textbf{(4.7.11)}$$
$$t = (-0.2029) \quad (2.1364)$$

اختبار Park يقترح أن هناك علاقة طردية لها معنوية إحصائية بين مربعات البواقي والمبيعات.

اختبار Glejser :

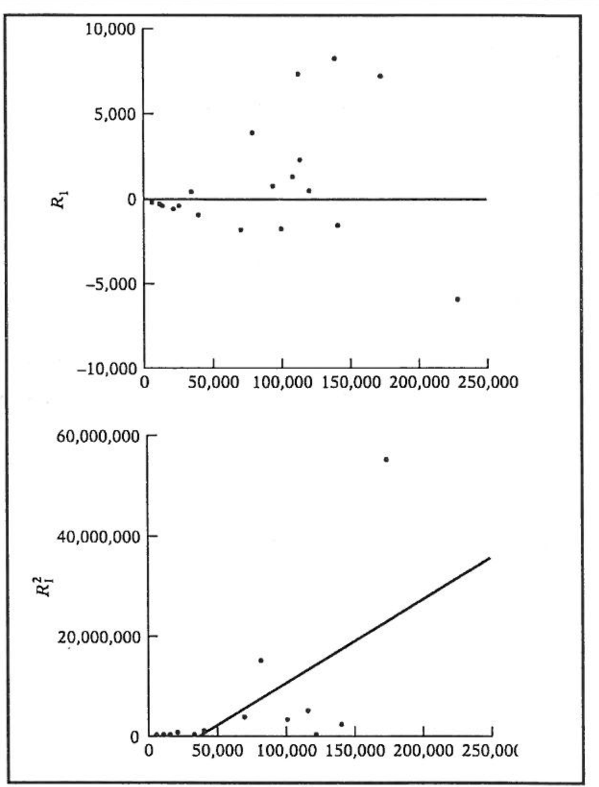
$$\left|\widehat{\hat{u}_i}\right| = 578.5710 + 0.0119\ \text{Sales}_i$$
$$\text{se} = (678.6950) \quad (0.0057) \quad r^2 = 0.214 \qquad \textbf{(5.7.11)}$$
$$t = (0.8524) \quad (2.0877)$$

اختبار Glejser يقترح أيضًا وجود علاقة منتظمة بين القيم المطلقة للبواقي والمبيعات، مما يزيد من إقبال أن يعاني انحدار (3.7.11) من اختلاف التباين.

جدول (5.11) الابتكار في أمريكا : نفقات البحث والتطوير (R & D) في الولايات المتحدة ، 1988 (وحدة كل الأشكال البيانية المليون دولار)

Innovation in America: Research and development (R & P) Expenditure in the unitcd states, 1988 (All Figures in Millions of Dollars)

Industry grouping	Sales	R&D expenses	Profits
1. Containers and packaging	6,375.3	62.5	185.1
2. Nonbank financial	11,626.4	92.9	1,569.5
3. Service industries	14,655.1	178.3	276.8
4. Metals and mining	21,869.2	258.4	2,828.1
5. Housing and construction	26,408.3	494.7	225.9
6. General manufacturing	32,405.6	1,083.0	3,751.9
7. Leisure time industries	35,107.7	1,620.6	2,884.1
8. Paper and forest products	40,295.4	421.7	4,645.7
9. Food	70,761.6	509.2	5,036.4
10. Health care	80,552.8	6,620.1	13,869.9
11. Aerospace	95,294.0	3,918.6	4,487.8
12. Consumer products	101,314.1	1,595.3	10,278.9
13. Electrical and electronics	116,141.3	6,107.5	8,787.3
14. Chemicals	122,315.7	4,454.1	16,438.8
15. Conglomerates	141,649.9	3,163.8	9,761.4
16. Office equipment and computers	175,025.8	13,210.7	19,774.5
17. Fuel	230,614.5	1,703.8	22,626.6
18. Automotive	293,543.0	9,528.2	18,415.4



شكل (13.11) البواقي R_1 ومربعات البواقي (R_1^2) على المبيعات

Residuals R_1 and Squared residuals (R_1^2) on Sales

اختبار White :

$$\widehat{|\hat{u}_i^2|} = -6{,}219{,}665 + 229.3508\ \text{Sales}_i - 0.000537\ \text{Sales}_i^2 \qquad (6.7.11)$$
$$\text{se} = (6{,}459{,}809) \quad (126.2197) \quad (0.0004)$$
$$t = (0.9628) \quad (1.8170) \quad (-1.3425)$$
$$R^2 = 0.2895$$

باستخدام قيمة R^2 و $n = 18$، نحصل على $n\,R^2 = 5.2124$ وهذه العينة تحت صحة الفرض العدمي والقائل بعدم وجود اختلاف في التباين، لها توزيع كاى التربيعي بدرجات حرية تساوي 2 [حيث إن لدينا متغيرين منحدرين في (6.7.11)] . قيمة p-value المرتبطة بقيمة كاى التربيعي المساوية لـ 5.2124 أو أكثر تساوى تقريباً 0.074 هذه القيمة للـ p-value تعتبر منخفضة، وذلك يجعل اختبار White يقترح وجود اختلاف في التباين.

في النهاية ووفقًا للأشكال البيانية الخاصة بالبواقي واختبارات Park، Glejser و White، فإن انحدار R&D (3.7.11) يبدو أنه يعاني من اختلاف التباين. وبما أن تباين الأخطاء الحقيقي غير معلوم، فإننا لن نستطيع استخدام طريقة المربعات الصغرى

المرجحة للحصول على الأخطاء القياسية المعدلة وفقًا لاختلاف التباين وقيم t. وبالتالي يجب أن نقوم ببعض التخمينات العلمية الخاصة بطبيعة تباين الأخطاء.

بالنظر إلى الأشكال البيانية الخاصة بالبواقي في (13.11)، يبدو أن تباين الخطأ يتناسب مع المبيعات كما في المعادلة (7.6.11)، أي أنه يمكن استخدام تحويلة الجذر التربيعي. ووفقًا لهذه التحويلة سنحصل على النتائج التالية:

$$\frac{\widehat{R\&D}}{\sqrt{Sales}} = -246.6769\frac{1}{\sqrt{Sales_i}} + 0.0367\sqrt{Sales_i}$$

$$\begin{aligned} se &= (381.1285) \quad (0.0071) \quad R^2 = 0.3648 \\ t &= (-0.6472) \quad (5.1690) \end{aligned} \tag{7.7.11}$$

من الممكن أن نضرب المعادلة السابقة في salesj (إذا رغبت في ذلك) حتى نعود مرة أخرى إلى الشكل الأصلي. بمقارنة (7.7.11) مع (3.7.11)، يمكنك أن تلاحظ أن معاملات الميل في المعادلتين تقريبًا متساوية ولكن أخطاؤها القياسية مختلفة. في (3.7.11) كانت تساوي 0.0083 في حين في (7.7.11) تساوي 0.0071 فقط. أي انخفضت بمقدار 14%للوصول إلى الاستنتاج النهائي من ذلك المثال دعنا نستعرض الأخطاء القياسية المتسقة لـ white في حالة اختلاف التباين كما سبق وناقشناها في الفقرة 6.11.

$$\widehat{R\&D}_i = 192.9931 + 0.0319\ Sales_i$$

$$\begin{aligned} se &= (533.9931) \quad (0.0101) \quad r^2 = 0.4783 \\ t &= (0.3614) \quad (3.1584) \end{aligned} \tag{8.7.11}$$

بمقارنة ذلك مع الانحدار الأصلي (3.7.11) (أي بدون التعديل الخاص باختلاف التباين) نجد أنه على الرغم من أن مقدرات المعالم لم تتغير (كما نتوقع) إلا أن الأخطاء القياسية لمعامل الجزء المقطوع من المحور الصادي قد انخفضت والأخطاء القياسية الخاصة بمعامل الميل قد زادت بمقدار بسيط. ولكن يجب أن نتذكر أن طريقة White خاصة بشكل محدد بالعينات كبيرة الحجم، ولدينا في هذا المثال 18 مفردة فقط.

8.11 تحذير من المبالغة في التعامل مع اختلاف التباين: A CAUTION ABOUT OVERREACTING TO HETEROSCEDASTICITY

بالعودة إلى مثال R&D المناقش في الفقرة السابقة، نرى أنه عندما استخدمنا تحويلة الجذر التربيعي لتصحيح اختلاف التباين الموجود في النموذج الأصلي (3.7.11)، فإن الأخطاء القياسية لمعامل الميل انخفضت، وقيم t المصاحبة لها زادت. هل هذا التغير يعتبر معنويًا ومهمًا بحيث يجب الاهتمام به عند التطبيق العملي؟ أو

بشكل آخر، متى يجب فعلاً أن نهتم بوجود مشكلة اختلاف التباين؟ أحد المؤلفين قال: " اختلاف التباين لم يكن يومًا السبب في تجاهل نموذج جيد ". [41]

هنا نجد أنه من الضروري الوضع في الاعتبار التحذير الذي صاغه John Fox كالتالي:

...... عدم تساوي تباين الأخطاء يعتبر مشكلة تستحق التعديل إذا كانت خطيرة فقط. فأثر عدم ثبات تباين الأخطاء على كفاءة مقدر المربعات الصغرى وعلى صحة الاستدلال باستخدام المربعات الصغرى يعتمد على عوامل عديدة، منها حجم العينة، درجة تغير σ_i^2، التعرف على قيم X [أي المنحدرة] والعلاقة بين تباين الخطأ و X's. وبالتالي فإنه يصعب التوصل إلى استنتاج عام خاص فقط بالمشكلة والضرر الذي ينتج عن اختلاف التباين. [42]

بالعودة إلى النموذج (1.3.11)، رأينا من قبل أن تباين مقدر الميل، $\text{var}(\hat{\beta}_2)$ معطى بالمعادلة التقليدية الموجودة في (3.2.11). وفقًا للـ GLS فإن تباين مقدر الميل، $\text{var}(\hat{\beta}_2^*)$ معطي في المعادلة (9.3.11)، ونعرف أن الأخير أكثر كفاءة من السابق. ولكن ما هو المدى الذي يستحق الاهتمام والذي يزيد فيه الأول (تباين OLS) عن الأخير (تباين GLS)؟

كقاعدة عامة، Fox يقترح أننا نبدأ بالقلق والاهتمام بهذه المشكلة " .... عندما يكون تباين الخطأ الأكبر يساوي 10 مرات التباين الأصغر " [43] وبالتالي بالعودة إلى محاكاة Monte carlo والتي سبق تقديمها في Mackinnon و Davidson، واذا استخدمنا $\alpha = 2$.

فإن تباين β_2 المقدرة هو 0.04 وفقًا للـ OLS، ويساوي 0.012 وفقًا للـ GLS، النسبة بين الأول والأخير حوالي 3.33. [44]

(41) N. Gregory Mankiw, "A Quick Refresher Course in Macroeconomics," Journal of Economic Literature, vol. XXVIII, December 1990, 9. 1648.

(42) John Fox, Applied Regression Analysis, Linear Models, and Related Methods, Sage Publications, California, 1997, p. 306.

(43) Ibid., p. 307.

(44) لاحظ أن لدينا مربعات الأخطاء القياسية والتي نحصل منها على التباين.

وبالتالي وفقًا لقاعدة Fox ، فإن حدة اختلاف التباين في هذه الحالة ليست شديدة ولا خطيرة بالدرجة التي تستحق القلق والاهتمام بوجود مثل هذه المشكلة.

تذكر أيضًا، بغض النظر عن اختلاف التباين، مقدرات OLS هي مقدرات خطية غير متحيزة وتقاربية (تحت مجموعة من الشروط العامة) وتؤول إلى التوزيع الطبقي (أي في أحجام العينات الكبيرة). وكما سنرى عندما نناقش مخالفات أخرى لفروض نموذج الانحدار الخطي التقليدي، فإن الاحتياطات الواجب استخدامها في هذه الفقرة تعتبر قاعدة عامة يمكن الاعتماد عليها.

9.11 التلخيص والاستنتاجات: SUMMARY AND CONCLUSIONS

1 - يفترض نموذج الانحدار الخطي التقليدي أن تكون الأخطاء u_i لها نفس التباين، σ^2. إذا لم يتحقق هذا الفرض تصبح لدينا مشكلة اختلاف التباين.

2 - اختلاف التباين لا يدمر خواص عدم التحيز والاتساق الخاصة بمقدرات OLS.

3 - ولكن هذه المقدرات لا تصبح لها أقل تباين أي لا تصبح كفئًا، أي لا تصبح BLUE.

4 - المقدرات الـ BLUE يمكن الحصول عليها باستخدام المربعات الصغرى المرجحة بافتراض أن تباين الأخطاء المختلف، σ_i^2 معلوم.

5 - في حالة وجود اختلاف في التباين، فإن تباينات مقدرات الـ OLS لا يتم الحصول عليها من معادلات OLS التقليدية. ولكن إذا صممنا على استخدام معادلات OLS التقليدية، فإن اختبارات t و F المرتبطة بها ستكون غير سليمة، وتؤدي إلى الوقوع في أخطاء استنتاجية.

6 - توثيق عواقب اختلاف التباين أسهل من اكتشافها، فعلى الرغم من وجود العديد من الاختبارات المتاحة للتعرف على الظاهرة، إلا أنه لا يوجد ما يحدد أيًا من هذه الاختبارات يكون مناسبًا وفقًا للبيانات محل الدراسة.

7 - حتى إذا افترضنا إمكانية اكتشاف وتحديد اختلاف التباين، فإنه ليس من السهل علاج هذه المسألة إذا كان حجم العينة كبيرًا، من الممكن أن تستخدم مقدرات الـ

OLS للأخطاء القياسية المصححة لـ White في حالة وجود اختلاف في التباين، ثم نقوم بعمل الاستدلال الإحصائي وفقًا لهذه الأخطاء القياسية.

8 - بخلاف ذلك، ووفقًا لبواقي الـ OLS، من الممكن أن نقوم بعمل تخمين علمي للنمط المتوقع لاختلاف التباين، ونحول البيانات الأصلية بطريقة تجعل البيانات المحولة لا تعاني من مشكلة اختلاف التباين.

تمارين : EXERCISES

أسئلة Questions

1.11 اشرح مع تعليل مختصر، أي من العبارات الآتية صحيح وأيها خطأ وأيها غير متأكد من إجابتك فيها.

(a) في حالة وجود اختلاف في التباين، فإن مقدرات OLS متحيزة وغير كفء.

(b) إذا كان هناك اختلاف في التباين، فإن اختبارات t و F غير صالحة للاستخدام.

(c) في حالة وجود اختلاف في التباين، فإن طريقة OLS العادية دائمًا تقدر الأخطاء القياسية بأعلى من قيمتها العقلية.

(d) إذا كانت البواقي المقدرة من انحدار OLS يظهر فيها نمط منتظم، فإن ذلك يعنى البيانات محل الدراسة يوجد فيها اختلاف في التباين.

(e) لا يوجد اختبار عام لاختلاف التباين خال من فرض خاص بالمتغير الذي يفترض أن يكون مرتبطًا مع حد الخطأ.

(f) إذا كان هناك خطأ في توصيف نموذج الانحدار (أي مثلاً تم حذف متغير مهم من النموذج)، فإن بواقي OLS ستظهر بنمط عشوائي.

(g) اذا تم حذف متغير منحدر من النموذج (عن طريق الخطأ) وهذا المتغير له تباين غير ثابت، فإن بواقي OLS سيكون فيها اختلاف في التباين.

2.11 في انحدار خاص بمتوسط الأجر (W, $) على عدد العمالة ($N$) لعينة عشوائية من 30 منشأة، تم الحصول على نتائج الانحدار التالية: (*)

(*) انظر في Dominick Salvatore, Managerial Economics, McGraw-Hill, New York, 1989, p. 157.

$$\widehat{W} = 7.5 + 0.009N \quad (1)$$
$$t = \text{n.a.} \quad (16.10) \qquad R^2 = 0.90$$

$$\widehat{W}/N = 0.008 + 7.8(1/N) \quad (2)$$
$$t = (14.43) \quad (76.58) \qquad R^2 = 0.99$$

(a) كيف يمكن تفسير نتائج الانحدارين السابقين؟

(b) ما الذي يفترضه الباحث حتى ينتقل من المعادلة (1) إلى (2)؟ هل كان يضع اختلاف التباين في الاعتبار؟ كيف عرفت ذلك؟

(c) هل يمكن أن تجد علاقة بين الميل والجزء المقطوع من المحور الصادي في النموذجين السابقين؟

(d) هل يمكنك مقارنة قيمة R^2 في النموذجين السابقين؟ علل إجابتك.

3.11 (a) هل يمكنك تقدير معالم النماذج التالية:

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i} + v_i$$

$$|\hat{u}_i| = \sqrt{\beta_1 + \beta_2 X_i^2} + v_i$$

باستخدام طريقة المربعات الصغرى العادية؟ علل إجابتك.

(b) إذا كان ذلك غير ممكن، هل يمكن أن تقترح طريقة، سواء رسمية أو غير رسمية، لتقدير معالم تلك النماذج؟ (انظر الفصل 14).

4.11 على الرغم من أن النماذج اللوغاريتمية الموضحة في المعادلة (12.6.11) دائمًا ما تقلل اختلاف التباين، إلا أن على الباحث أن يتعامل بمنتهى الحرص مع خصائص مقدر الخطأ الخاص بتلك النماذج. فمثلاً للنموذج:

$$Y_i = \beta_1 X_i^{\beta_2} u_i \quad (1)$$

يمكن أن يكتب على الصورة التالية:

$$\ln Y_i = \ln \beta_1 + \beta 2 \ln X_i + \ln u_i$$

(a) إذا كان $\ln u_i$ له توقع يساوي الصفر، ما هو التوزيع المتوقع لـu_i؟

(b) إذا كان $E(u_i) = 1$ هل يعني ذلك أن $E(\ln u_i) = 0$؟ علل إجابتك.

(c) إذا كان $E(\ln u_i)$ لا يساوي الصفر، ما الذي يمكن عمله لجعله يساوي الصفر؟

5.11 اثبت أن β_2^* الموجود في (8.3.11) يمكن كتابته على الصورة:

$$\beta_2^* = \frac{\sum w_i y_i^* x_i^*}{\sum w_i x_i^{2*}}$$

و $\text{var}(\beta_2^*)$ الموجود في (9.3.11) يمكن كتابته أيضًا على الصورة:

$$\text{var}(\beta_2^*) = \frac{1}{\sum w_i x_i^{2*}}$$

حيث $y_i^* = Y_i - \bar{Y}^*$ ، $x_i^* = X_i - \bar{X}^*$ تمثل الانحرافات على الأوساط المرجحة $\bar{Y}^*$ و $\bar{X}^*$ المعرفة كالتالي:

$$\bar{Y}^* = \sum w_i Y_i \Big/ \sum w_i$$

$$\bar{X}^* = \sum w_i X_i \Big/ \sum w_i$$

6.11 لأغراض بيداغوجيا (علم أصول التدريس)، قام كل من Hanushek و Jackson لتقدير النموذج التالي:

$$C_t = \beta_1 + \beta_2 \text{GNP}_t + \beta_3 \text{D}_t + u_i \qquad (1)$$

حيث C_t = نفقات الاستهلاك الخاص التجميعية في العام t.

GNP_t = الناتج المحلي الإجمالي في العام t. و D = نفقات الدفاع المحلية في العام t، الهدف من التحليل هو دراسة أثر نفقات الدفاع على النفقات الأخرى في الاقتصاد.

اذا افترضنا أن $\sigma_t^2 = \sigma^2(\text{GNP}_t)^2$ ، فإن (1) تتحول إلى المعادلة (2) ونقدر التالي:

$$C_t/\text{GNP}_t = \beta_1\ (1/\text{GNP}_t) + \beta_2 + \beta_3\ (D_t/\text{GNP}_t) + u_t/\text{GNP}_t \qquad (2)$$

النتائج العملية للبيانات من الفترة 1946 إلى 1975 كانت كالتالي (الأخطاء القياسية معطاة بين الأقواس)(*):

$$\hat{C}_t = 26.19 \quad + 0.6248\ \text{GNP}_t - \ 0.4398\ D_t$$
$$(2.73) \qquad (0.0060) \qquad (0.0736) \qquad R^2 = 0.999$$

$$\widehat{C_t/\text{GNP}_t} = 25.92\,(1/\text{GNP}_t) + 0.6246 \quad - \ 0.4315\,(D_t/\text{GNP}_t)$$
$$(2.22) \qquad (0.0068) \qquad (0.0597) \qquad R^2 = 0.875$$

(a) ما هي الفروض التي فرضها الباحثون حول طبيعة اختلاف التباين؟ هل يمكنك تفسيرها؟

(*) Eric A. Hanushek and John E. Jackson, Statistical Methods for Social Scientists, Academic, New York, 1977, p. 160.

(b) قارن نتائج الانحدارين . هل التحويلة المستخدمة على النموذج الأصلي حسنت النتائج ، أي هل التحويلة قللت من الأخطاء القياسية؟ علل إجابتك .

(c) هل يمكنك مقارنة قيم R^2، علل إجابتك . (لاحظ التالي : جاوب على هذا السؤال من خلال اختيارك للمتغيرات التابعة) .

7.11 بالرجوع إلى الانحدارات المقدرة (2.6.11) و(3.6.11) . هل نتائج الانحدار متقاربة . ما هو السبب في ذلك؟

8.11 أثبت أنه إذا كانت $w_i = w$ أي ثانية ، لكل i ، فإن β_2^* و $\hat{\beta}_2$ وأخطاؤهما القياسية تكون متساوية .

9.11 بالرجوع إلى المعادلات (2.2.11) و (3.2.11)، افترض أن

$$\sigma_i^2 = \sigma^2 k_i$$

حيث σ^2 ثابت و k_i أوزان محددة، وليست بالضرورة متساوية . وفقًا لهذا الفرض ، اثبت أن التباين المعطى في (2.2.11) يمكن كتابته كالتالي :

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \cdot \frac{\sum x_i^2 k_i}{\sum x_i^2}$$

المقدار الأول في الجزء الموجود في الجانب الأيمن من المعادلة يمثل معادلة التباين الموجودة في (3.2.11)، أي أن $\operatorname{var}(\hat{\beta}_2)$ ثابت التباين . ما الذي يمكن قوله عن طبيعة العلاقة بين $\operatorname{var}(\hat{\beta}_2)$ في حالة اختلاف التباين ، وفي حالة ثبات التباين؟ (ملاحظة : اختير المقدار الثاني الموجود على الجانب الأيمن من المعادلة السابقة) . هل يمكنك التوصل إلى أي استنتاجات عن العلاقة بين (2.2.11) و(3.2.11)؟

10.11 اعتبر النموذج التالي :

$Y_i = \beta_2 Y_i + u_i$ (لاحظ أنه لا يوجد جزء ثابت مقطوع من المحور الصادي)

إذا علمت أن $\operatorname{var}(u_i) = \sigma^2 X_i^2$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2 \sum X_i^4}{\left(\sum X_i^2\right)^2}$$

مسائل : Problems

11.11 استخدم البيانات المعطاة في جدول 1.11 وقم بعمل انحدار لمتوسط التعويضات Y على متوسط الإنتاج X، واستخدم حجم العمالة كوحدة المشاهدات . فسر النتائج وحدد ما إذا كانت متفقة مع نتائج (3.5.11) .

(a) من الانحدار السابق احصل على البواقي $\hat{u}_i$.

(b) وفقًا لاختبار Park، قم بعمل انحدار لـ $\ln \hat{u}_i^2$ على $\ln X_i$ واوجد انحدار (4.5.11).

(c) وفقًا لطريقة Glejser، قم بعمل انحدار لـ $|\hat{u}_i|$ و $\sqrt{X_i}$ ثم قم بعمل انحدار وعلق على النتائج.

(d) اوجد معامل ارتباط الرتب بين $|\hat{u}_i|$ و X_i وعلق على طبيعة اختلاف التباين إذا وجدت في البيانات محل الدراسة.

12.11 جدول (6.11) يعطي بيانات خاصة بنسبة المبيعات إلى المبيعات نقدًا في مصانع الولايات المتحدة مقسمة وفقًا لحجم الأصول في فترة التكوين للفترة من 1971-I إلى 1973-IV (البيانات ربع سنوية). نسبة المبيعات إلى المبيعات نقدًا قد ينظر إليها كمقياس لسرعة الدخل في القطاع الخدمي أي عدد مرات عودة الدولار.

(a) لكل حجم من أحجام الأصول، احسب الوسط والانحراف المعياري لنسبة المبيعات إلى المبيعات نقدًا.

(b) ارسم القيمة المتوقعة ضد الانحراف المعياري المسحوبات في **a** باستخدام حجم الأصول كوحدة المشاهدات.

(c) عن طريق المتوسطات الخاصة بالانحدار المناسب، حدد ما إذا كان الانحراف المعياري للنسبة يتزايد مع العينة المتوسطة أم لا. إذا كانت الإجابة بلا، كيف يمكنك تفسير ذلك؟

جدول (6.11) حجم الأصول (بالمليون دولار) Asset size (Millions of dollars)

Year and quarter	1–10	10–25	25–50	50–100	100–250	250–1000	1000 +
1971–I	6.696	6.929	6.858	6.966	7.819	7.557	7.860
–II	6.826	7.311	7.299	7.081	7.907	7.685	7.351
–III	6.338	7.035	7.082	7.145	7.691	7.309	7.088
–IV	6.272	6.265	6.874	6.485	6.778	7.120	6.765
1972–I	6.692	6.236	7.101	7.060	7.104	7.584	6.717
–II	6.818	7.010	7.719	7.009	8.064	7.457	7.280
–III	6.783	6.934	7.182	6.923	7.784	7.142	6.619
–IV	6.779	6.988	6.531	7.146	7.279	6.928	6.919
1973–I	7.291	7.428	7.272	7.571	7.583	7.053	6.630
–II	7.766	9.071	7.818	8.692	8.608	7.571	6.805
–III	7.733	8.357	8.090	8.357	7.680	7.654	6.772
–IV	8.316	7.621	7.766	7.867	7.666	7.380	7.072

المصدر: التقرير المالي الربع سنوي للمؤسسة الصناعية، وزارة التجارة الدولية ووزارة التبادل الاقتصادي، الحكومة الأمريكية، أعداد متنوعة (محسوبة)

(d) إذا كانت هناك علاقة إحصائية معنوية بين الاثنين. كيف يمكنك تحويل البيانات بحيث لا يوجد فيها اختلاف التباين؟

13.11 اختبار Bartlett لثبات التباين(*). افترض أن لدينا k من تباينات العينة المستقلة s_1^2، s_2^2، . . . ، s_k^2 بدرجات الحرية f_1، f_2، . . . ، f_k كل منها يتبع توزيعًا طبيعيًا متوسطه μ وتباينه σ_i^2. افترض أيضًا أننا نريد اختيار الفرض العدمي $H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2 = \sigma^2$، أي أن كل تباين عينة يعتبر مقدرًا لنفس تباين المجتمع σ^2.

تحت صحة الفرض العدمي، فإن:

$$s^2 = \frac{\sum_{i=1}^{k} f_i s_i^2}{\sum f_i} = \frac{\sum f_i s_i^2}{f}$$

يعتبر مقدرًا (تجميعيًا) مشتركًا لتباين المجتمع σ^2، حيث $f_i = (n_i - 1)$، n_i عدد المشاهدات في المجموعة i وحيث $f = \sum_{i=1}^{k} f_i$.

أوضح Bartlett أن هذا الفرض العدمي يمكن اختياره بالنسبة A/B والتي تؤول تقاربيًا إلى توزيع كاي التربيعي بدرجات حرية $k-1$ حيث

$$A = f \ln s^2 - \sum (f_i \ln s_i^2)$$

و

$$B = 1 + \frac{1}{3(k-1)}\left[\sum\left(\frac{1}{f_i}\right) - \frac{1}{f}\right]$$

طبق اختبار Bartlett على بيانات جدول 1.11 واثبت أن فرض تساوي بيانات مجتمعات تعويضات العمال وفقًا لكل حجم من أحجام العمالة في المنشآت المختلفة لا يمكن رفضه عند مستوى معنوية 5% لاحظ أن: f_i، درجات الحرية لكل عينة تباين، يساوي 9، حيث إن n_i لكل عينة (عينة من العمال) تساوي 10.

14.11 اعتبر نموذج الانحدار التالي المار بنقطة الأصل:

$$Y_i = \beta X_i + u_i, \quad \text{for } i = 1, 2$$

إذا علمت أن $u_1 \sim N(0,\sigma^2)$ و $u_2 \sim N(0,2\sigma^2)$ وأنهما مستقلان إحصائيًا. إذا كانت $X_1 = +1$ و $X_2 = -1$، احصل على مقدار المربعات الصغرى المرجح (WLS) لـ B وتباينه. إذا فترضت بشكل غير سليم أن تباينات الخطأين متساويان (أي

(*) انظر "Properties of Sufficiency and Statistical Tests," Proceedings of the Royal Society of London A, vol. 160, 1937, p. 268.

يساوين (σ^2)، ما هو مقدر OLS لـβ؟ وتباينه؟ قارن هذه المقدرات مع نظيرها المحسوبة باستخدام WLS؟ ما هو استنتاجك العام؟(*)

15.11 جدول (7.11) يعطي بيانات خاصة بـ 81 سيارة، هذه البيانات تشتمل على MPG (متوسط الأميال بالنسبةللجالون)، HP (قوة المحرك) VOL (السعة بالـ feet المكعب)، SP (السرعة القصوى بالميل في الساعة) و WT (وزن السيارة بالـ Ib 100).

جدول (7.11) بيانات بالأميال عن سيارة الراكب **Passenger Car Milage Data**

Observation	MPG	SP	HP	VOL	WT	Observation	MPG	SP	HP	VOL	WT
1	65.4	96	49	89	17.5	42	32.2	106	95	106	30.0
2	56.0	97	55	92	20.0	43	32.2	109	102	92	30.0
3	55.9	97	55	92	20.0	44	32.2	106	95	88	30.0
4	49.0	105	70	92	20.0	45	31.5	105	93	102	30.0
5	46.5	96	53	92	20.0	46	31.5	108	100	99	30.0
6	46.2	105	70	89	20.0	47	31.4	108	100	111	30.0
7	45.4	97	55	92	20.0	48	31.4	107	98	103	30.0
8	59.2	98	62	50	22.5	49	31.2	120	130	86	30.0
9	53.3	98	62	50	22.5	50	33.7	109	115	101	35.0
10	43.4	107	80	94	22.5	51	32.6	109	115	101	35.0
11	41.1	103	73	89	22.5	52	31.3	109	115	101	35.0
12	40.9	113	92	50	22.5	53	31.3	109	115	124	35.0
13	40.9	113	92	99	22.5	54	30.4	133	180	113	35.0
14	40.4	103	73	89	22.5	55	28.9	125	160	113	35.0
15	39.6	100	66	89	22.5	56	28.0	115	130	124	35.0
16	39.3	103	73	89	22.5	57	28.0	102	96	92	35.0
17	38.9	106	78	91	22.5	58	28.0	109	115	101	35.0
18	38.8	113	92	50	22.5	59	28.0	104	100	94	35.0
19	38.2	106	78	91	22.5	60	28.0	105	100	115	35.0
20	42.2	109	90	103	25.0	61	27.7	120	145	111	35.0
21	40.9	110	92	99	25.0	62	25.6	107	120	116	40.0
22	40.7	101	74	107	25.0	63	25.3	114	140	131	40.0
23	40.0	111	95	101	25.0	64	23.9	114	140	123	40.0
24	39.3	105	81	96	25.0	65	23.6	117	150	121	40.0
25	38.8	111	95	89	25.0	66	23.6	122	165	50	40.0
26	38.4	110	92	50	25.0	67	23.6	122	165	114	40.0
27	38.4	110	92	117	25.0	68	23.6	122	165	127	40.0
28	38.4	110	92	99	25.0	69	23.6	122	165	123	40.0
29	46.9	90	52	104	27.5	70	23.5	148	245	112	40.0
30	36.3	112	103	107	27.5	71	23.4	160	280	50	40.0
31	36.1	103	84	114	27.5	72	23.4	121	162	135	40.0
32	36.1	103	84	101	27.5	73	23.1	121	162	132	40.0
33	35.4	111	102	97	27.5	74	22.9	110	140	160	45.0
34	35.3	111	102	113	27.5	75	22.9	110	140	129	45.0
35	35.1	102	81	101	27.5	76	19.5	121	175	129	45.0
36	35.1	106	90	98	27.5	77	18.1	165	322	50	45.0
37	35.0	106	90	88	27.5	78	17.2	140	238	115	45.0
38	33.2	109	102	86	30.0	79	17.0	147	263	50	45.0
39	32.9	109	102	86	30.0	80	16.7	157	295	119	45.0
40	32.3	120	130	92	30.0	81	13.2	130	236	107	55.0
41	32.2	106	95	113	30.0						

لاحظ أن : VOL = السعة بالـ feet المكعب .

HP = قوة المحرك .

(*) مـأخـوذة من F. A. F. Seber, Linear Regression Analysis, John Wiley & Sons, New York, 1977, p. 64.

MPG = متوسط الأميال لكل جالون
SP = السرعة القصوى بالميل في الساعة .
WT = وزن السيارة ، بالـ 100 Ib .
Observation = رقم السيارة المشاهدة (نوع السيارة غير مسجل) .

المصدر : U.S. Environmental Protection Agency, 1991, Report EPA/AA/CTAB/91-02.

(a) اعتبر النموذج التالي :

$$MPG_i = \beta_1 + \beta_2 SP + \beta_3 HP + \beta_4 WT + u_i$$

قدر معالم هذا النموذج وفسر النتائج . هل لها أي معنى اقتصادي؟

(b) هل تتوقع أن يكون هناك اختلاف في التباين الخاص بالأخطاء في النموذج السابق؟

(c) استخدم اختبار White لاكتشاف ما إذا كان تباين الخطأ يوجد فيه اختلاف في التباين .

(d) احصل على قيم الأخطاء القياسية و t المتسقة لـ White في حالة اختلاف التباين ، وقارن نتائجك مع نظيرها الذي نحصل عليه من الـ OLS .

(e) إذا كان هناك اختلاف في التباين . كيف يمكنك تحويل البيانات ، بحيث إن البيانات المحولة يكون لها تباين أخطاء ثابتًا؟ وضح الخطوات الحسابية الضرورية لعمل ذلك .

16.11 نفقات الغذاء في الهند . في جدول (8.2) لدينا بيانات عن نفقات الغذاء والنفقات الإجمالية لـ 55 أسرة في الهند .

(a) قم بعمل انحدار لنفقات الغذاء على النفقات الإجمالية واختبر البواقي التي ستحصل عليها من هذا الانحدار .

(b) ارسم البواقي التي ستحصل عليها من (a) ضد النفقات الإجمالية . هل ترى أي نمط محدد؟

(c) إذا كان الرسم البياني في (b) يظهر وجود اختلاف في التباين ، طبق اختبارات Park و Glejser و White . لترى ما إذا كان الانطباع الذي توصلت إليه في (b) عن اختلاف التباين تؤيده هذه الاختبارات أم لا .

(d) احصل على الأخطاء القياسية المتسقة لـ White في حالة اختلاف التباين وتاريخها مع الأخطاء القياسية للـ OLS . حدد ما إذا كان من الضروري تصحيح اختلاف التباين في هذا المثال أم لا .

17.11 أعد حل تمرين 16.11 ولكن في هذه المرة قم بعمل انحدار للوغاريتم نفقات الغذاء على لوغاريتم النفقات الإجمالية. إذا لاحظت اختلافًا في التباين من النموذج الخطي لتمرين 16.11 ولكن لم تلاحظه في النموذج اللوغاريتمي الخطي، ما الاستنتاج الذي يمكن التوصل إليه؟ وضح الخطوات الحسابية الضرورية لذلك.

18.11 طريقة مختصرة لاختبار White. كما سبق وذكرنا، اختبار White يستهلك درجات حرية عديدة إذا كانت هناك متغيرات منحدرة كثيرة، وبالتالي إذا استخدمنا كل المتغيرات المنحدرة وقيمها المربعة وحاصل ضربها لن تقوم بتقدير الانحدار مثل الموجود في (22.5.11)، لماذا لا تقوم ببساطة بتقدير الانحدار التالي:

$$\hat{u}_i^2 = \alpha_1 + \alpha_2 \hat{Y}_i + \alpha_2 \hat{Y}_i^2 + \nu_i$$

حيث $\hat{Y}_i$ هي القيمة المقدرة لـ Y (أي المتغير المنحدر عليه) والتي نحصل عليها من أي نموذج مقدر؟ ففي النهاية، $\hat{Y}_i$ هي متوسط مرجح للمتغيرات المنحدرة ومعاملات الانحدار المقدرة لمثل الأوزان.

احصل على R^2 من الانحدار السابق واستخدم (22.5.11) لاختيار فرض عدم وجود اختلاف في التباين.

طبق الاختبار السابق لمثال نفقات الغذاء الموجود في تمرين 16.11.

19.11 بالرجوع إلى مثال R&D الموجود في الفقرة 7.11. قم بإعادة حل المثال باستخدام الأرباح كمتغير منحدر. مسبقًا هل تتوقع أن تكون مختلفة عن النتائج التي حصلنا عليها عندما استخدمنا المبيعات كمتغير منحدر؟ علل إجابتك.

جدول (8.11) وسيط رواتب أساتذة الإحصاء، 2000 – 2001

Median Salaries of full professors in Statistics, 2000-2001

Years in rank	Count	Median
0 to 1	11	$69,000
2 to 3	20	$70,500
4 to 5	26	$74,050
6 to 7	33	$82,600
8 to 9	18	$91,439
10 to 11	26	$83,127
12 to 13	31	$84,700
14 to 15	15	$82,601
16 to 17	22	$93,286
18 to 19	23	$90,400
20 to 21	13	$98,200

تابع - جدول (8.11) وسيط رواتب أساتذة الإحصاء ، 2000 - 2001

Years in rank	Count	Median
22 to 24	29	$100,000
25 to 27	22	$99,662
28 to 32	22	$116,012
33 or more	11	$85,200

المصـــدر: American Statistical Association, "2000–2001 Salary Report of Academic Statisticians, "Amstat News, Issue 282. December 2000, p. 4.

20.11 جدول (8.11) يعطي بيانات عن وسيط رواتب أساتذة في علم الإحصاء في جامعات بحثية في الولايات المتحدة عن السنة الدراسية 2000 – 2001 .

(a) ارسم وسيط الرواتب ضد السنوات المرتبة (كمقياس لعدد سنوات الخبرة). حتى تستطيع الرسم، افترض أن الوسيط هو القيمة المتوسطة للسنوات المرتبة. فمثلاً الراتب المساوي لـ 74,050$ في المدى 4 – 5 يشار إليه بـ 4.5 سنة مرتبة، وهكذا.

للمجموعة الأخيرة، افترض أن المدى من 33 إلى 35 .

(b) اعتبر نماذج الانحدار التالية :

$$Y_i = \alpha_1 + \alpha_2 X_i + u_i \quad (1)$$

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + v_i \quad (2)$$

حيث Y = وسيط الراتب، X = السنة المرتبة (مقاسة بوسيط المدى) و u و v مقادير الأخطاء. هل يمكنك تفسير لماذا قد يفضل النموذج (2) عن النموذج (1)؟ من البيانات المعطاة قدر كل من النموذجين.

(c) إذا وجد اختلاف في التباين في النموذج (1) ولم يوجد في النموذج (2)، ما الاستنتاج الذي يمكنك التوصل إليه؟ وضح الخطوات الضرورية.

(d) إذا لاحظنا وجود اختلاف في التباين في النموذج (2)، كيف يمكنك تحويل البيانات بحيث لا يصبح اختلاف التباين موجوداً؟

21.11 إذا أعطيت البيانات التالية :

RSS_1 وفقًا لأول 30 مفردة = 55، درجات حرية = 25

RSS_2 وفقًا لآخر 30 مفردة = 140 ، درجات حرية = 25

قم بعمل اختبار Goldfeld–Quandt لاختلاف التباين عند مستوى معنوية 5% .

22.11 جدول (9.11) يعطي بيانات عن نسبة التغير السنوية لأسعار الأسهم (Y) وسعر المستهلك (X) لبيانات مقطعية من 20 دولة.

(a) ارسم البيانات في شكل انتشار.

(b) قم بعمل انحدار لـ Y على X واختبر البواقي التي ستحصل عليها من النموذج. ماذا تلاحظ؟

(c) هل ترى أي قيم شاذة في بيانات دولة تشيلي؟ قم بعمل الانحدار مرة أخرى مع حذف بيانات دولة تشيلي. اختبر الآن بواقي نموذج الانحدار. ماذا تلاحظ؟

(d) بناء على نتائج الفقرة (b) نستنتج أن هناك اختلافًا في تباين الخطأ، ولكن ووفقًا لنتائج (c) نستنتج شيئًا آخر. ما الاستنتاج العام الذي يمكن أن نتوصل إليه؟

جدول (9.11) أسعار الأسهم والاستهلاك – فترة ما بعد الحرب العالمية الثانية (حتى 1969)

Stock and Consumer prices - post world warII period (through 1969)

	Rate of change, % per year	
Country	Stock prices, Y	Consumer prices, X
1. Australia	5.0	4.3
2. Austria	11.1	4.6
3. Belgium	3.2	2.4
4. Canada	7.9	2.4
5. Chile	25.5	26.4
6. Denmark	3.8	4.2
7. Finland	11.1	5.5
8. France	9.9	4.7
9. Germany	13.3	2.2
10. India	1.5	4.0
11. Ireland	6.4	4.0
12. Israel	8.9	8.4
13. Italy	8.1	3.3
14. Japan	13.5	4.7
15. Mexico	4.7	5.2
16. Netherlands	7.5	3.6
17. New Zealand	4.7	3.6
18. Sweden	8.0	4.0
19. United Kingdom	7.5	3.9
20. United States	9.0	2.1

المصدر: Phillip Cagan, Common Stock Values and Inflation: The Historical Record of Many Countries, National Bureau of Economic Research, Suppl., March 1974, Table 1, p. 4.

Appendix 11A ملحق A 11

1.A11 إثبات المعادلة (2.2.11) :

من ملحق A3 ، الفقرة 3.A3 نجد أن :

$$\operatorname{var}(\hat\beta_2) = E(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2 + 2 \text{ cross-product terms})$$
$$= E(k_1^2u_1^2 + k_2^2u_2^2 + \cdots + k_n^2u_n^2)$$

بما أن توقعات مقادير حاصل الضرب تساوي الصفر وفقًا لفرض عدم وجود ارتباط تسلسلي ، نجد أن :

$$\operatorname{var}(\hat\beta_2) = k_1^2E(u_1^2) + k_2^2E(u_2^2) + \cdots + k_n^2E(u_n^2)$$

حيث إن k_i معلومة . (لماذا؟)

$$\operatorname{var}(\hat\beta_2) = k_1^2\sigma_1^2 + k_2^2\sigma_2^2 + \cdots + k_n^2\sigma_n^2$$

حيث $E(u_i^2) = \sigma_i^2$

$$\operatorname{var}(\hat\beta_2) = \sum k_i^2\sigma_i^2$$
$$= \sum\left[\left(\frac{x_i}{\sum x_i^2}\right)^2\sigma_i^2\right] \qquad \text{since } k_i = \frac{x_i}{\sum x_i^2} \qquad (2.2.11)$$
$$= \frac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2}$$

2.A11 طريقة المربعات الصغرى المرجحة : The Method of Weighted Least Squares

لشرح هذه الطريقة ، تستخدم نموذج انحدار ثنائي المتغيرات $Y_i = \beta_1 + \beta_3 X_i = u_i$

طريقة المربعات الصغري غير المرجحة تعتمد على تصغير التالي :

$$\sum \hat u_i^2 = \sum (Y_i - \hat\beta_1 - \hat\beta_2 X_i)^2 \qquad (1)$$

للحصول على المقدرات ، أما طريقة المربعات الصغرى المرجحة فتعتمد على تصغير مجموع مربعات البواقي المرجحة كالتالي :

$$\sum w_i\hat u_i^2 = \sum w_i(Y_i - \hat\beta_1^* - \hat\beta_2^* X_i)^2 \qquad (2)$$

حيث β_1^* و β_2^* هي مقدرات المربعات الصغرى المرجحة وأوزان الترجيح w_i هي كالتالي :

$$w_i = \frac{1}{\sigma_i^2} \qquad (3)$$

أي أن الأوزان تتناسب عكسيًا مع تباين u_i أو Y_i مشروطة بالـ X_i حيث إنه من المعلوم أن $\text{var}(u_i|X_i) = \text{var}(Y_i|X_i) = \sigma_i^2$

عندما نقوم بعمل تفاضل لـ (2) بالنسبة β_1^* و$\hat{\beta}_2^*$، نحصل على:

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_1^*} = 2\sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-1)$$

$$\frac{\partial \sum w_i \hat{u}_i^2}{\partial \beta_2^*} = 2\sum w_i (Y_i - \hat{\beta}_1^* - \hat{\beta}_2^* X_i)(-X_i)$$

وبمساواة السابق بالصفر، نحصل على المعادلتين الطبيعيتين التاليتين:

$$\sum w_i Y_i = \hat{\beta}_1^* \sum w_i + \hat{\beta}_2^* \sum w_i X_i \qquad (4)$$

$$\sum w_i X_i Y_i = \hat{\beta}_1^* \sum w_i X_i + \hat{\beta}_2^* \sum w_i X_i^2 \qquad (5)$$

لاحظ التشابه بين هذه المعادلات الطبيعية ونظيرها الخاص بالمربعات الصغرى غير المرجحة.

عند حل هذه المعادلات آنيًا، نحصل على:

$$\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^* \qquad (6)$$

و

$$\hat{\beta}_2^* = \frac{(\sum w_i)(\sum w_i X_i Y_i) - (\sum w_i X_i)(\sum w_i Y_i)}{(\sum w_i)(\sum w_i X_i^2) - (\sum w_i X_i)^2} \qquad (8.3.11) = (7)$$

تباين $\hat{\beta}_2^*$ المعطى في (9.3.11) يمكن الحصول عليه كما سبق وأوضحنا في ملحق A3، الفقرة 3.A3.

لاحظ أن: $\bar{Y}^* = \sum w_i Y_i / \sum w_i$ و $\bar{X}^* = \sum w_i X_i / \sum w_i$. ومن السهل إثبات أن متوسطات هذه الأوزان تتساوى مع متوسطات $\bar{Y}$ و $\bar{X}$ العادية أو غير المرجحة عندما تكون $w_i = w$ أي ثابتة لكل قيم الـ i.

3.A11 إثبات أن $E(\hat{\sigma}^2) \neq \sigma^2$ في حالة وجود اختلاف في التباين:

Proof that $E(\hat{\sigma}^2) \neq \sigma^2$ in the presence of Heteroscedasticity

اعتبر نموذج الانحدار التالي ثنائي المتغيرات:

$$Y_i = \beta_1 + \beta_2 X_i + u_i \qquad (1)$$

حيث $\text{var}(u_i) = \sigma_i^2$

الآن

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{n-2} = \frac{\sum (Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum [\beta_1 + \beta_2 X_i + u_i - \hat{\beta}_1 - \hat{\beta}_2 X_i]^2}{n-2}$$

$$= \frac{\sum [-(\hat{\beta}_1 - \beta_1) - (\hat{\beta}_2 - \beta_2) X_i + u_i]^2}{n-2} \tag{2}$$

لاحظ أن $(\hat{\beta}_1 - \beta_1) = -(\hat{\beta}_2 - \beta_2)\bar{X} + \bar{u}$ ، وبالتعويض عن ذلك في (2) وادخال التوقع على طرفي المعادلة نحصل على:

$$E(\hat{\sigma}^2) = \frac{1}{n-2}\left\{-\sum x_i^2 \text{var}(\hat{\beta}_2) + E\left[\sum (u_i - \bar{u})^2\right]\right\}$$

$$= \frac{1}{n-2}\left[-\frac{\sum x_i^2 \sigma_i^2}{\sum x_i^2} + \frac{(n-1)\sum \sigma_i^2}{n}\right] \tag{3}$$

حيث استخدمنا (2.2.11) .

تستطيع أن ترى من (3)، إذا كان هناك ثبات في التباين، أي إذا كان $\sigma_i^2 = \sigma^2$ لكل i، $E(\hat{\sigma}) = \sigma^2$. وبالتالي القيمة المتوقعة لـ $\hat{\sigma}^2$ المحسوبة $= \sum \hat{u}^2 / (n-2)$ لن تكون مساوية للقيمة σ^2 الحقيقية في حالة وجود اختلاف في التباين[1].

4.A11 الأخطاء القياسية اللامعلمية (الثابتة) لـ White:
White's Robust Standard Errors

لاستقراء الأخطاء القياسية المعدلة لـ White في حالة اختلاف التباين، دعنا نعتبر نموذج الانحدار التالي ثنائي المتغيرات:

$$Y_i = \beta_1 + \beta_1 X_i + u_i \qquad \text{var}(u_i) = \sigma_i^2 \tag{1}$$

كما هو موضح في (2.2.11) فإن

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2} \tag{2}$$

حيث σ_i^2 لا تشاهد مباشرة، فإن white يقترح استخدام $\hat{u}_i^2$ القيمة المربعة للبواقي لكل i، بدلاً من σ_i^2 ونقدر $\text{var}(\hat{\beta}_2)$ كالتالي:

$$\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \hat{u}_i^2}{\left(\sum x_i^2\right)^2} \tag{3}$$

(1) للمزيد من التفصيل اطلع على Jan Kmenta, Elements of Econometrics, 2d. ed., Macmillan, New York, 1986, pp. 276–278.

أوضح white أن (3) يعتبر مقدراً متسقاً لـ (2)، أي أنه مع زيادة حجم العينة، فإن (3) تؤول إلى (2).[2]

لاحظ أنه إذا لم تشتمل الحزمة الإحصائية على طريقة الأخطاء القياسية الثابتة لـ white، يمكنك القيام بها كما هو موضح في (3) عن طريق أولاً إجراء انحدار OLS العادي، والحصول على بواقي هذا الانحدار ثم استخدام المعادلة (3).

طريقة White يمكن تقييمها إلى نموذج انحدار يشتمل على k متغير كالتالى :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i \tag{4}$$

تباين أي من معاملات الانحدار الجزيئية، مثل $\hat{\beta}_j$، نحصل عليه كالتالي :

$$\operatorname{var}(\hat{\beta}_j) = \frac{\sum \hat{w}_{ji}^2 \hat{u}_i^2}{\left(\sum \hat{w}_{ji}^2\right)^2} \tag{5}$$

حيث $\hat{u}_i$ هي البواقي التي نحصل عليها من نموذج الانحدار (الأصلي)(4) و $\hat{w}_j$ هي البواقي التي نحصل عليها من الانحدار (المساعد) للمتغيرات المنحدرة X_j على باقي المتغيرات المنحدرة في (4).

من الواضح أن ذلك يمثل طريقة مستهلكة للوقت لتقدير (5) لكل متغير X. بالطبع يمكن تجنب كل ذلك إذا كانت الحزمة الإحصائية المستخدمة يمكنها القيام بذلك بشكل أوتوماتيكي. الحزم الإحصائية مثل Pc Give، Eviews، Microfit، Shazam، Stata و Limdep يمكن استخدامها للحصول بسهولة على الأخطاء القياسية الثابتة لـ white في حالة اختلاف التباين.

---

(2) لنكون أكثر دقة، فإن n مضروبة في (3) تؤول احتماليًا إلى $E[(X_i - \mu x)^2 u_i^2]/(\sigma_x^2)^2$، وهذا يساوي النهاية الاحتمالية لـ n مضروب في (2) حيث n تمثل حجم العينة، μ_x القيمة المتوقعة لـ X و σ_x^2 تباين (مجتمع) X. لمزيد من التفاصيل، انظر في M. Wooldridge, Introductory Econometrics; A Modern Approach, South-Western Publishing, 2000, p. 250.

الفصل الثاني عشر

الارتباط الذاتي: ماذا يحدث إذا كانت حدود الخطأ مرتبطة؟

AUTOCORRELATION: WHAT HAPPENS IF THE ERROR TERMS ARE CORRELATED

دعنا نتذكر أن هناك بوجه عام ثلاثة أنواع مختلفة من البيانات المتاحة للتحليل العملي: (1) البيانات المقطعية، (2) السلاسل الزمنية و(3) مزيج من البيانات المقطعية وبيانات السلاسل الزمنية والمعروفة باسم البيانات التجميعية. عندما استعرضنا نموذج الانحدار الخطي التقليدي (CLRM) في الجزء 1، افترضنا بعض الفروض التي ناقشناها في الجزء 1.7. عمومًا لاحظنا أيضًا أن كل الفروض ليست بالضرورة متحققة لكل أنواع البيانات. وعلى وجه الخصوص، فقد رأينا في الفصل السابق، أن فرض ثبات التباين، أو تساوي تباين الأخطاء، قد لا يتحقق دائمًا في حالة البيانات المقطعية. بمعنى آخر دائمًا ما تعاني البيانات المقطعية من مشكلة اختلاف التباين.

عمومًا في الدراسات التي تعتمد على بيانات مقطعية، عادة ما يتم تجميع البيانات على أساس عينة عشوائية من الوحدات المقطعية، مثل الأسرة (مثل تحليل دالة الاستهلاك) أو الوحدة المنتجة (مثل تحليل دراسة الاستثمار)، وبالتالي لا يوجد سبب مسبق يجعلنا نعتقد أن مقدار الخطأ الخاص بالأسرة أو الوحدة المنتجة قد يكون مرتبطًا مع أسرة أخرى أو وحدة منتجة أخرى. وإذا وجد هذا الارتباط عن طريق المصادفة في وحدات البيانات المقطعية، فإن ذلك يسمى ارتباطًا مزيفًا، أي أنه ارتباط في الفراغ أكثر منه في التحليل محل الدراسة. عمومًا من المهم أن نتذكر أنه في تحليل البيانات المقطعية، ترتيب البيانات لابد أن يكون ذا مغزى معين، أو معنى اقتصادي محدد، وذلك يساعدنا في معرفة ما إذا كان الارتباط (المزيف) موجودًا أم لا.

الوضع كله سيتغير تمامًا إذا كنا نتعامل مع بيانات سلاسل زمنية. حيث إن هذه المشاهدات تتبع ترتيبًا طبيعيًا عبر الزمن، وبالتالي فالمفردات عادة ما تكون مترابطة، خصوصًا إذا كان الزمن بين المفردات بعضها وبعض يعتبر زمنًا قصيراً، مثل يوم أو أسبوع، شهر أو حتى سنة. فاذا كنت مهتمًا بدراسة مؤشرات أسعار الأسهم مثل مؤشر S&P 500 أو مؤشر Dow jones اليومي، فإنه من المعتاد أن تجد هذه المؤشرات ترتفع وتنخفض لأيام عديدة. وبالتالي في موقف مثل ذلك فرض عدم وجود ارتباط ذاتي أو تسلسلي في حد الخطأ (وهذا الفرض هو فرض خاص بـ CLRM) لن يتحقق.

في هذا الفصل، فإننا نتناول بعمق هذا الفرض، ونحاول الإجابة عن الأسئلة التالية:

1 - ما هي طبيعة الارتباط الذاتي؟

2 - ما هي العواقب النظرية والعملية للارتباط الذاتي؟

3 - حيث إن فرض عدم وجود ارتباط ذاتي خاص بحد الخطأ u_t، كيف يمكن للفرد أن يعرف ما إذا كان هناك ارتباط ذاتي أم لا؟ لاحظ أن t تعني أننا نتعامل مع بيانات سلاسل زمنية.

4 - كيف يمكن معالجة مشكلة الارتباط الذاتي؟

سيجد القارئ أن هذا الفصل شبيه بالفصل السابق الخاص باختلاف التباين. حيث إنه مع افتراض وجود كل من اختلاف التباين والارتباط الذاتي، فإن مقدرات الـ OLS العادية، الخطية ستكون غير متحيزة وتؤول تقاربيًا إلى التوزيع الطبيعي (أي في حالة أحجام العينات الكبيرة).[1] إلا أنها لن تكون ذات التباين الأقل من بين فئة المقدرات الخطية غير المتحيزة. باختصار، لن تكون كفئًا مقارنةً للمقدرات الأخرى الخطية غير المتحيزة. بعبارة أخرى، لن يكونوا BLUE. وبالتالي فإن اختبارات t، F، X^2 التقليدية قد لا يجوز استخدامها.

(1) لمزيد من التفاصيل، انظر William H. Greene, Econometric Analysis, 4th ed., Prentice Hall, N.J.,2000, Chap. 11, and Paul A. Rudd, An Introduction to Classical Econometric Theory, Oxford University Press, 2000, Chap. 19.

1.12 طبيعة المشكلة : THE NATURE OF THE PROBLEM

مصطلح «الارتباط الذاتي» يمكن أن يُعرف على أنه «ارتباط بين عناصر السلسلة الواحدة من المشاهدات المرتبة زمنيًا [كما في بيانات السلاسل الزمنية] أو مكانيًا [كما في البيانات المقطعية]» .[2] في إطار موضوع الانحدار، فإن نموذج الانحدار الخطي التقليدي يفترض أن مثل هذا الارتباط الذاتي لا يتواجد في حد الخطأ u_i، رمزيًا . .

$$E(u_i u_j) = 0 \qquad i \neq j \qquad (5.2.3)$$

أي ببساطة، النموذج التقليدي يفترض أن حد الخطأ الخاص بأي مفردة لا يتأثر بحد الخطأ الخاص بأي مفردة أخرى. فمثلاً إذا كنا نتعامل مع بيانات ربع سنوية خاصة بانحدار الناتج على مدخل العمالة ورأس المال وإذا كان مثلاً، هناك تغيير مفاجئ في العمالة يؤثر على الناتج في الربع الأول من السنة لا يوجد ما يجعلنا نعتقد أن هذا التغيير سيظهر مرة أخرى في الربع التالي. بمعني أنه إذا كان هناك انخفاض في الربع التالي. وبالمثل إذا كنا نتعامل مع بيانات مقطعية خاصة بانحدار نفقات استهلاك الأسرة على دخلها إثر زيادة دخل أسرة ما على استهلاكها لا يتوقع أن يؤثر على نفقات استهلاك أسرة أخرى.

وبالتالي، إذا وجد عدم الاستقلال بهذا الشكل، فإن لدينا ارتباطًا ذاتيًا ورمزيًا يمكن التعبير عنه كالتالي:

$$E(u_i u_j) \neq 0 \qquad i \neq j \qquad (1.1.12)$$

في هذه الحالة، التغيير الحادث نتيجة التغيير المفاجئ في هذا الربع قد يؤثر بشكل كبير على ناتج الربع التالي، والزيادة الحادثة في استهلاك إحدى الأسر ممكن أن تؤثر على أسرة أخرى وتزيد من نفقات استهلاكها.

وقبل أن نبدأ معرفة لماذا يوجد هذا الارتباط الذاتي، فإنه من الضروري أن نوضح بعض الأمور الخاصة بالمصطلحات الإحصائية المتعلقة بهذه النقطة، فعلى الرغم من شيوع استخدام الارتباط الذاتي، والارتباط التسلسلي كمترادفين، إلا أن بعض

(2) Maurice G. Kendall and William R. Buckland, A Dictionary of Statistical Terms, Hafner Publishing Company, New York, 1971, 9. 8.

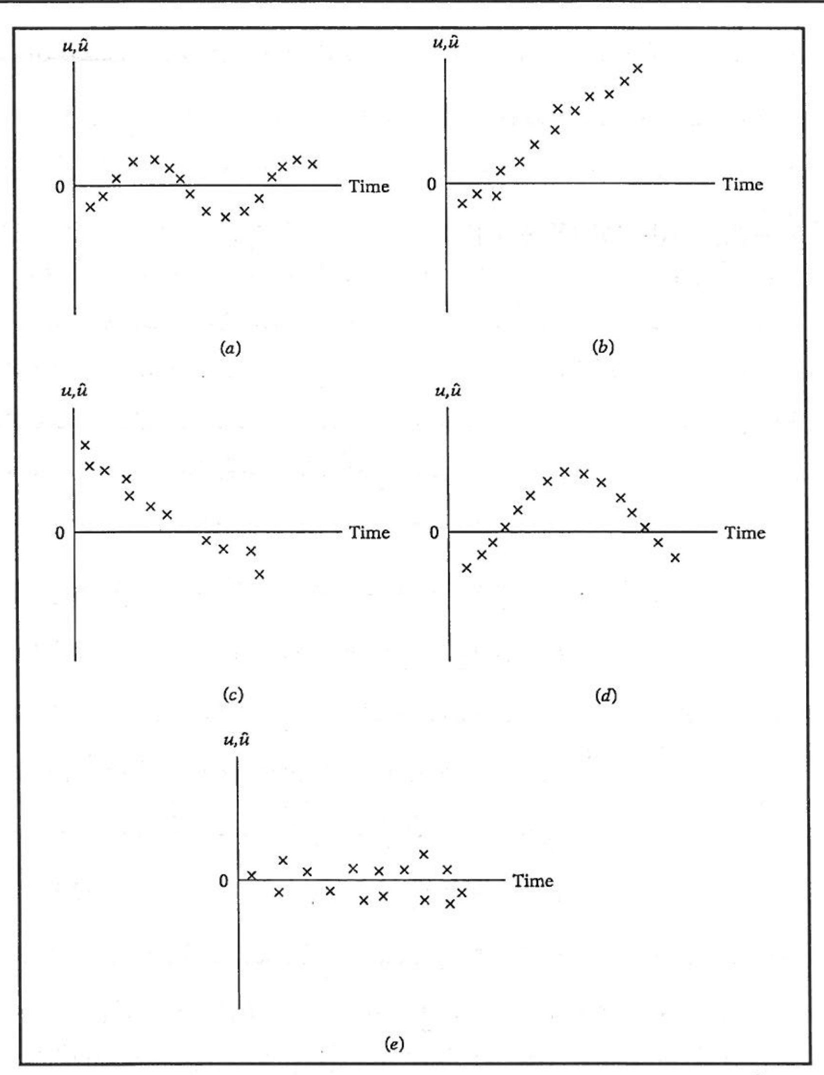
الباحثين يفضلون التفرقة بين المصطلحين. فمثلاً، Tintner يُعرِّف الارتباط الذاتي بأنه «ارتباط بين مشاهدات سلسلة معينة عند الفجوات الزمنية المختلفة، معبر عنها بوحدات زمنية» في حين أنه يُعرِّف مصطلح الارتباط التسلسلي على أنه «ارتباط بين متسلسلتين مختلفتين وفقًا للفجوات الزمنية»[3] وبالتالي الارتباط بين سلسلتين مثل u_1، u_2، . . . ، u_{10} و u_2، u_3، . . . ، u_{11} بحيث إن الأولى هي الأخيرة. ولكن بعد إزاحتها فترة زمنية واحدة يُعتبر ارتباطًا ذاتيًا في حين أن الارتباط بين المتسلسلتين مثل $u1$، u_2، . . . ، u_{10} وv_2، v_3، . . . ، v_{11} حيث u، v سلسلتين زمنيتين مختلفتين تمامًا يسمى ارتباطًا تسلسليًا. وعلى الرغم من أهمية التفرقة بين المصطلحين، إلا أننا في هذا الكتاب سنقوم باستخدام المصطلحين كمترادفين.

دعنا نستعرض بيانيًا بعض الأنماط الخاصة بمجالات وجود وعدم وجود ارتباط ذاتي، وهذه الأنماط معطاة في شكل (1.12). الأشكال (a1.12 إلى d) توضح الأنماط المختلفة الممكن تواجدها بين u's . شكل (a1.12) يوضح النمط الدوري، شكل (b1.12 و c) يوضح الاتجاه العام الخطي المتزايد والمتناقص في حد الخطأ، أما شكل (d1.12) يوضح الاتجاه العام الخطي والتربيعي الموجودان في حد الخطأ. أما شكل (e1.12) فهو الشكل الوحيد الذي لا يمثل أي نمط منتظم، مما يدعم فرض وجود ارتباط ذاتي، والخاص بنموذج الانحدار الخطي التقليدي.

السؤال الطبيعي الآن هو: لماذا يوجد هذا الارتباط التسلسلي؟ هناك العديد من الأسباب، ودعنا نذكر فيما يلي بعضًا منها:

القصور الذاتي: الخاصية الساكنة في معظم السلاسل الزمنية الاقتصادية هي القصور الذاتي أو الركود. فمثلاً كما هو معروف، سلسلة مثل GNP مؤشرات الأسعار، الإنتاج، العمالة والبطالة يوجد فيها شكل دوري.

(3) Gerhard Tintner, Econometrics, John Wiley & Sons, New York, 1965.



شكل (1.12) أنماط تمثل وجود وعدم وجود ارتباط ذاتي Patterns of autocorrelation and nonautocorrelation

إذا بدأنا من قاع فترة الفتور الاقتصادي، وعندما بدأ التحسن التدريجي للاقتصاد، معظم السلاسل الزمنية بدأت في الصعود. وفي هذا التصاعد تكون قيمة السلسلة عند نقطة زمنية ما أكبر من قيمتها السابقة في الزمن. مما يعني أنها تكون قيمًا عظمى ويظل الوضع كما هو حتى يحدث شيء ما (مثل زيادة في سعر الفائدة أو الضرائب أو كليهما معًا) وتبدأ السلسلة في الانخفاض مرة أخرى. ولذلك فإن في الانحدار الخاص ببيانات السلاسل الزمنية تكون القيم المتتالية في الزمن للمشاهدات مترابطة ذاتيًا.

تحيز التوصيف.. حالة استبعاد متغيرات: Specification Bias.. Excluded variables Case

في التحليل الفعلي، يقوم الباحث في البداية بعمل نموذج انحدار، والذي قد لا يكون «الأفضل» على الإطلاق. وبعد تحليل الانحدار، يبدأ الباحث بعمل توافق بين النتائج وما كان متصور الحصول عليه مسبقًا. إذا لم يحدث مثل هذا التوافق فلابد من البدء في العملية الجراحية. مثلاً، قد يقوم الباحث بعمل رسم بياني للبواقي $\hat{u}_i$ التي حصل عليها من الانحدار المقدر، ويلاحظ نمطًا محدداً كالموجود في الشكل (a1.12) إلى (b1.12). هذه البواقي (والتي تعبر عن الـ u_i) قد تجعل الباحث يقترح وجود بعض المتغيرات المهمة ولكنها غير موجودة في النموذج بضرورة إدخالها في نموذج الانحدار. هذه الحالة تعتبر تحيزًا في التوصيف راجع إلى استبعاد بعض المتغيرات. وعادة ما يؤدي إدخال هذه المتغيرات إلى إلغاء نمط الارتباط المشاهد بين البواقي.

وعلى سبيل المثال، دعنا نفترض أن لدينا نموذج الطلب التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t \qquad (2.1.12)$$

حيث Y = الكمية المطلوبة من اللحوم البقري، X_2 = سعر اللحوم البقري، X_3 = دخل المستهلك، X_4 = سعر لحم الخنزير و t = الزمن.[4]

عمومًا لبعض الأسباب سنقوم بعمل الانحدار التالي:

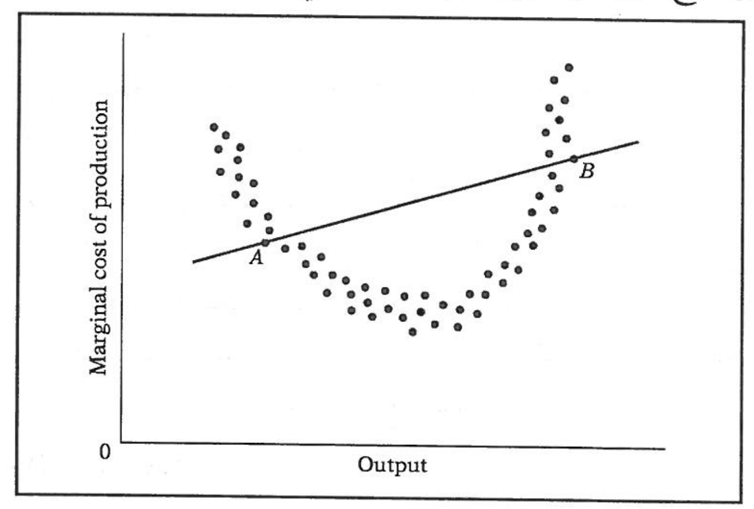
$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + v_t \qquad (3.1.12)$$

إذا كان (2.1.12) هو النموذج «السليم» أو «الحقيقي»، فإن القيام بالانحدار (3.1.12) يعني أن $v_t = \beta_4 X_{4t} = u_t$. ونستطيع أن نقول إن سعر لحم الخنزير يؤثر على استهلاك اللحم البقري، وبالتالي فإن حد الخطأ v سيظهر هذا النمط المنتظم، وبالتالي نحصل على الارتباط الذاتي (المزيف). اختبار بسيط لذلك هو إجراء كل من (2.1.12) و(3.1.12) ونرى ما إذا كان هناك ارتباط ذاتي، إذا وجد، ظاهر في النموذج (3.1.12) ومختفي في النموذج (2.1.12) عند تطبيقه.[5]

(4) للتوضيح سنستخدم الترميز (t) للتعبير عن بيانات السلاسل الزمنية والترميز المعتاد (i) للتعبير عن البيانات المقطعية.

(5) إذا وجدت أن المشكلة الحقيقية هي ناتجة عن تحيز التوصيف، وليس الارتباط الذاتي، بالتالي كما سنرى لاحقًا في (الفصل 13)، فإن مقدرات OLS لمعلمات (13.12) قد تكون متحيزة وغير متسقة أيضًا.

الطريقة الفعلية لاكتشاف الارتباط الذاتي، سنناقشها بالتفصيل في الفقرة 6.12، والخاصة بالرسوم البيانية لبواقي كل من الانحدارين (2.1.12) و(3.1.12). وسنلقي الضوء بوضوح على ما يسمى بالارتباط التسلسلي.



شكل (2.12) تحيز التوصيف : الشكل الدالي الخاطئ Specification bias: Incorrect Functional form

تحيز التوصيف.. الشكل الدالي الخاطئ: Specification bias: Incorrect Functional form

افترض أن النموذج الصحيح أو«الحقيقي» في دراسة خاصة بالناتج والتكلفة له الشكل التالي:

$$\text{Marginal cost}_i = \beta_1 + \beta_2 \text{ output}_i + \beta_3 \text{ output}_i^2 + u_i \qquad \textbf{(4.1.12)}$$

ولكننا قمنا بتقدير النموذج التالي:

$$\text{Marginal cost}_i = \alpha_1 + \alpha_2 \text{ output}_i + v_i \qquad \textbf{(5.1.12)}$$

منحنى التكلفة الحدية الخاص بالنموذج «الحقيقي» موضحة في الشكل (2.12) ومعه أيضًا منحنى التكلفة الخطية الخاص بالنموذج «غير الحقيقي». كما نرى من الشكل (2.12) بين النقاط A وB منحنى التكلفة الحدية الخطي سيقدر بشكل متسق قيمة للتكلفة الحدية الحقيقية بأعلى من قيمتها العقلية. يجب أن نتوقع مثل هذه النتيجة، لأن مقدار الخطأ v_i سيظهر ارتباطًا ذاتيًا كنتيجة لاستخدام شكل دالي خاطئ. في (الفصل 13) سنستعرض طرقًا عديدة لاكتشاف تحيز التوصيف.

ظاهرة نسيج العنكبوت: Cobweb Phenomenon

المعروض من السلع الزراعية يظهر ما يسمى بظاهرة نسيج العنكبوت، حيث إن المعروض يتأثر بالسعر لفترة زمنية سابقة، حيث إن قرارات الكمية المعروضة تأخذ وقتًا حتى تنفذ (فترة الحمل). وبالتالي في بداية التخطيط السنوي للمحصول، فإن المزارعين سيتأثرون بسعر العام السابق، وبالتالي تكون دالة العرض كالتالي:

$$\text{Supply}_t = \beta_1 + \beta_2 P_{t-1} + u_i \qquad (6.1.16)$$

افترض أنه عند نهاية الفترة t، السعر P_t أصبح أقل مما كان عليه في P_{t-1} وبالتالي فإنه في الفترة $t+1$ فإن المزارعين قد يقررون تقليل الكمية المزروعة عن التي زرعوها بالفعل في الفترة t. ويتضح في هذه الحالة أن الخطأ u_t لا يتوقع أن يكون عشوائيًا، حيث إنه إذا كان هناك فائض مزروع في السنة t، فالأغلب سيتم تقليل الإنتاج في $t+1$ وهكذا يجعلنا ندخل في نمط النسيج العنكبوتي.

الفترات الفاضلة: Lags

في انحدار السلاسل الزمنية الخاص بنفقات الاستهلاك على الدخل، من المعتاد أن نجد نفقات الاستهلاك في الفترة الحالية تعتمد على نفقات الاستهلاك في الفترة السابقة (بالإضافة إلى عوامل أخرى). أي أن:

$$\text{Consumption}_t = \beta_1 + \beta_2 \text{ income}_t + \beta_3 \text{ consumption}_{t-1} + u_t \qquad (7.1.12)$$

انحدار مثل الموجود في (7.1.12) يعرف باسم **الانحدار الذاتي autoregnerssion** حيث إن أحد المتغيرات المفسرة هو المتغير التابع في نقطة زمنية سابقة (سندرس هذه النماذج في الفصل 17). المنطق وراء مثل هذا النموذج يمكن تفسيره ببساطة. المستهلكون لا يغيرون عاداتهم الاستهلاكية وذلك بسبب العديد من الأسباب الفنية والنفسية والمجتمعية. فإذا تجاهلنا المقدار الخاص بالفترة الزمنية السابقة الموجود في (7.1.12) فسيعكس مقدار الخطأ نمطًا منتظمًا نتيجة تأثير استهلاك الفترة الزمنية السابقة على الاستهلاك في الفترة الحالية.

«معالجة» البيانات: "Manipulation" of Data

في التحليل التطبيقي الفعلي، تكون البيانات الخام عادة ما تم «معالجتها». فعلى سبيل المثال، في انحدار السلاسل الزمنية والخاص بالبيانات الربع سنوية، عادة

ما تكون هذه البيانات تم الحصول عليها من بيانات شهرية، عن طريق جمع بيانات الشهور الثلاثة ثم قسمة المجموع على 3. هذه المتوسطات تمهد البيانات، فتعالج التغيرات الشهرية للبيانات. وبالتالي الرسم البياني الخاص بالبيانات الربع سنوية يبدو ممهدًا أكثر من البيانات الشهرية. وهذا التمهيد نفسه قد يؤدي إلى ظهور نمط منتظم في الخطأ، مما يظهر الارتباط الذاتي. مصدر آخر من مصادر المعالجة هو إقحام أو إخراج بعض البيانات.

فعلى سبيل المثال، تعداد السكان يتم كل 10 سنوات في بلد ما، الأخير كان في 2000 والسابق له كان في 1990. الآن إذا كانت هناك حاجة للحصول على بيانات لسنوات معينة في الفترة ما بين 1990 و2000، التصرف الأكثر شيوعًا هو إدماجهما معًا وفقًا لشروط الإضافة والحذف. كل هذه الأساليب من التعامل مع البيانات قد تؤدي إلى وجود نمط منتظم قد لا يتواجد في البيانات الأصلية.(6)

تحويل البيانات: Data Transformation

كمثال لذلك، دعنا نعتبر النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad \textbf{(8.1.12)}$$

حيث Y = نفقات الاستهلاك، X = الدخل، حيث (8.1.12)، تتحقق عند كل نقطة زمنية، فإنها متحققة أيضًا في الفترة الزمنية السابقة $(t-1)$ وبالتالي يمكن أن نكتب (8.1.12) كالتالي:

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad \textbf{(9.1.12)}$$

Y_{t-1}، X_{t-1}، u_{t-1} تسمى بقيم الفترات السابقة بالترتيب، حيث الفترة الفاصلة هنا هي فترة واحدة فقط. سنرى أهمية قيم الفترات السابقة في هذا الفصل في العديد من الفقرات.

الآن إذا طرحنا (9.1.12) من (8.1.12) سنحصل على:

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \qquad \textbf{(10.1.12)}$$

(6) للمزيد من التفاصيل، انظر William H. Greene, op. cit., p. 526.

حيث Δ تعرف بمعامل الفروق الأولى حيث يجعلنا هذا المعامل نهتم بفروق المتغير. وبالتالي $\Delta Y_t = Y_t - Y_{t-1}$ و $\Delta X_t = X_t - X_{t-1}$ و $\Delta u_t = u_t - u_{t-1}$، لأهداف تطبيقه دعنا نكتب (10.1.12) كالتالي:

$$\Delta Y_t = \beta_2 \Delta X_t + \Delta u_t \qquad (11.1.12)$$

حيث $v_t = \Delta u_t = u_t - u_{t-1}$

المعادلة (9.1.12) تعرف باسم الشكل الطبقي، والمعادلة (10.1.12) تعرف باسم شكل الفروق الأولى. كل منهما شائع الاستخدام في التحليل التطبيقي. فعلى سبيل المثال، إذا كانت Y و X الموجودة في (9.1.12) تمثل لوغاريتم نفقات الاستهلاك والدخل، فإن (10.1.12) الموجود فيها ΔY و ΔX ستمثل التغيير في لوغاريتم نفقات الاستهلاك والدخل. ولكن كما نعرف، التغير في لوغاريتم المتغير هو تغير نسبي أو تغير مئوي إذا تم ضرب السابق في 100. وبالتالي بدلاً من دراسة العلاقة بين المتغيرات على المستوى أو الشكل الطبقي نكون مهتمين بدراسة العلاقة على مستوى معدل النمو أو التغيير.

والآن إذا كان حد الخطأ الموجود في (8.1.12) مستوفيًا فروض OLS التقليدية وبالأخص فرض عدم وجود ارتباط ذاتي، يمكن إثبات أن حد الخطأ v_i الموجود في (11.1.12) يوجد فيه ارتباط ذاتي. (الإثبات موجود في الملحق A12، الفقرة 12 1.A). ونلاحظ هنا أن نماذج مثل (11.1.12) تعرف باسم نماذج الانحدار الديناميكية أي نماذج تشتمل على متغيرات منحدرة في فترات زمنية سابقة. سنتطرق لمثل هذه النماذج بالتفصيل في الفصل 17.

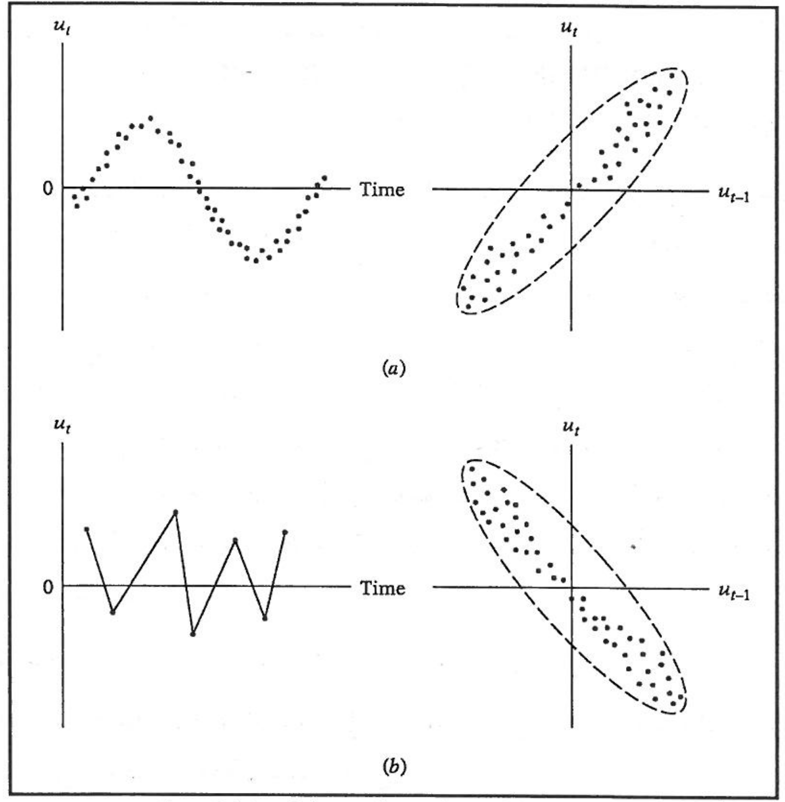
الفكرة في المثال السابق تعتمد على أنه أحيانًا يكون لدينا ارتباط ذاتي كنتيجة لعمل تحويلة ما في النموذج الأصلي.

عدم السكون: Nonstationarity

ذكرنا في (الفصل 1) أنه عند التعامل مع بيانات السلاسل الزمنية، يجب علينا أن نتحقق من سكون السلسلة الزمنية محل الدراسة. وعلى الرغم من أننا سنناقش موضوع عدم السكون في الفصول الخاصة بالاقتصاد القياسي المرتبط بالسلاسل الزمنية في الجزء V من الكتاب، إلا أنه بوجه عام، تعتبر السلسلة الزمنية ساكنة إذا

كانت خصائصها (مثل الوسط، التباين والتغاير) لا تتغير مع الزمن. أي ساكنة، إذا لم يتحقق ذلك، يكون لدينا سلسلة زمنية غير ساكنة.

كما سيتضح في الجزء V من الكتاب، في نموذج الانحدار مثل الموجود في (8.1.12) يكون من المحتمل أن نجد كلاً من X و Y غير ساكنين، وبالتالي يكون الخطأ u غير ساكن أيضًا.[7] في مثل هذه الحالة، سيكون هناك ارتباط ذاتي في حد الخطأ.



شكل (3.12) : (a) ارتباط ذاتي طردي و(b) ارتباط ذاتي عكسي

(a) Positive and (b) negative autocorrelation

وخلاصة ما سبق، أن هناك العديد من الأسباب التي قد تؤدي إلى جعل حد الخطأ في نموذج الانحدار يوجد فيه ارتباط ذاتي. في المتبقي من الفصل، سنستعرض بعض التفاصيل الخاصة بمشاكل الارتباط الذاتي، وما يمكن فعله لمواجهة مثل هذه

(7) كما سنرى في الجزء V ، حتى إذا كان X و Y غير ساكنين ، قد نجد أن u ساكنًا . وسنستعرض أثر ذلك لاحقًا .

المشاكل. ونلاحظ أيضًا أن الارتباط الذاتي قد يكون طرديًا كما في شكل (a3.12) وقد يكون عكسيًا أيضًا، إلا أن معظم السلاسل الزمنية الخاصة بالمتغيرات الاقتصادية عادة ما يظهر فيها ارتباط ذاتي طردي، حيث إن معظمهم يتحركون في شكل تصاعدي أو تنازلي عبر الزمن، ونادرًا ما نرى التحرك الثابت ارتفاعًا وانخفاضًا، كما في الشكل (b3.12).

2.12 تقديرات OLS في حالة وجود ارتباط ذاتي:
OLS ESTIMATION IN THE PRESENCE OF AUTOCORRELATION

ماذا يحدث لمقدرات OLS وتبايناتها إذا كان لدينا ارتباط ذاتي في حد الخطأ مع افتراض أن $E(u_t u_{t+s}) \neq 0$، $(s \neq 0)$ والاحتفاظ بباقي الفروض الخاصة بالنموذج التقليدي متحققة؟[8] لاحظ مجددًا أننا الآن نستخدم التغير t على حد الخطأ للتغيير عن التعامل مع بيانات سلاسل زمنية. دعنا نعود مرة أخرى إلى نموذج الانحدار الذي يشتمل على متغيرين اثنين لتوضيح فكرتنا الرئيسية، أي دعنا نستخدم $Y_t = \beta_1 + \beta_2 X_t + u_t$. دعنا أولاً نتفق على أن الآلية التي سيتم بها توليد u_t بحيث $E(u_t u_{t+s}) \neq 0$، $(s \neq 0)$ يجب أن تكون آلية معتمدة على فروض عامة غير مرتبطة بأي تطابق محدد. كنقطة بداية، وكأول تقريب يمكن أن نفترض أن حد الخطأ يمكن توليده وفقًا للتالي:

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \qquad (1.2.12)$$

حيث ρ (=rho) معروفة باسم معامل الارتباط الذاتي، و t هو حد الخطأ العشوائي الذي يستوفي فروض OLS التقليدية، وهي

$$E(\varepsilon_t) = 0$$

$$\operatorname{var}(\varepsilon_t) = \sigma_\varepsilon^2 \qquad (2.2.12)$$

$$\operatorname{cov}(\varepsilon_t, \varepsilon_{t+s}) = 0 \qquad s \neq 0$$

(8) اذا كان $s = 0$، يمكن أن نحصل على $E(u_t^2)$. فعندما يكون $E(u_t) = 0$ فرضيًا، فإن $E(u_t^2)$ ستمثل تباين حد الخطأ، والذي لن يساوي الصفر بالتأكيد (لماذا؟)

في أدبيات الهندسة، حد الخطأ السابق ووفقًا لخصائصه السابقة يسمى white noise error term . ما يعنيه (1.2.12) أن قيمة الخطأ عند النقطة الزمنية t يساوي rho مضروبة في قيمته عند النقطة الزمنية السابقة $t-1$ بالإضافة إلى حد عشوائي للخطأ.

العملية (1.2.12) تسمى عملية لـ Markov من الرتبة الأولى أو للاختصار عملية الارتباط الذاتي من الرتبة الأولى ويرمز لها **AR(1)**. المصطلح ارتباط ذاتي مناسب حيث إن (1.2.12) يمكن تفسيرها أن انحدار u_t على نفسه في فترة زمنية سابقة، ويعتبر من الرتبة الأولى لأن u_t وقيمة سابقة له بفترة زمنية واحدة هما الموجودان في النموذج، أي أن أقصى قيمة سابقة مستخدمة هي 1. إذا كان النموذج على الشكل $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$ سيكون لدينا **AR(2)** أو محلية ارتباط ذات من الرتبة الثانية وهكذا. سندرس العمليات من الرتب الأعلى في الفصول الخاصة بالسلاسل الزمنية الاقتصادية في الجزء V.

وبشكل عابر، لاحظ أن rho، معامل الارتباط الذاتي في (1.2.12) يمكن أن يفسر على أنه معامل الارتباط الذاتي من الدرجة الأولى أو بشكل أكثر دقة معامل الارتباط الذاتي عند Lag 1 .[9]

وفقًا لنموذج AR(1)، يمكن إثبات أن (انظر الملحق A12)، الفقرة 2.A12)

$$\operatorname{var}(u_t) = E(u_t^2) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{3.2.12}$$

$$\operatorname{cov}(u_t, u_{t+s}) = E(u_t u_{t-s}) = \rho^s \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{4.2.12}$$

$$\operatorname{cor}(u_t, u_{t+s}) = \rho^s \tag{5.2.12}$$

حيث $\operatorname{cov}(u_t, u_{t+s})$ يمثل التغاير بين حدود الخطأ المتباعدة s فترة زمنية و$\operatorname{cor}(u_t, u_{t+s})$ يمثل الارتباط بين حدود الخطأ المتباعدة s فترة زمنية. لاحظ أنه وفقًا

---

(9) يمكن تفسير ذلك بسهولة، فوفقًا للتعريف، معامل ارتباط (المجتمع) بين u_t و u_{t-1} هو

$$\rho = \frac{E\{[u_t - E(u_t)][u_{t-1} - E(u_{t-1})]\}}{\sqrt{\operatorname{var}(u_t)}\sqrt{\operatorname{var}(u_{t-1})}}$$

$$= \frac{E(u_t u_{t-1})}{\operatorname{var}(u_{t-1})}$$

وبما أن $E(u_t) = 0$ لكل t و $\operatorname{var}(u_t) = \operatorname{var}(u_{t-1})$ حيث إننا نفترض ثبات التباين. ويمكن للقارئ أن يتحقق من أن ρ تمثل معامل الميل في انحدار u_t على u_{t-1} .

لخاصية التماثل في التغاير والارتباط فإن $\text{cor}(u_t, u_{t-s}) = \text{cov}(u_t, u_{t+s})$ و $\text{cor}(u_t, u_{t-s}) = \text{cov}(u_t, u_{t+s})$.

وبما أن p ثابتة بين 1− و 1+، فإن (3.2.12) تثبت أنه وفقًا لـ AR(1) فإن تباين u_t مازال ثابتًا ولكن u_t مرتبطة ليس فقط بالقيمة السابقة لها في الزمن ولكن بالعديد من القيم السابقة في الزمن. ومن المهم ملاحظة أن $|\rho| < 1$، أي أن القيمة المطلقة لـ rho أقل من الواحد الصحيح. فمثلاً إذا كانت ρ تساوي 1، فإن التباينات والارتباطات المذكورة أعلى ستكون غير معرفة. أما إذا كان $|\rho| < 1$، فإننا نقول إن لدينا عملية AR (1) المتمثلة في (1.2.12) عملية ساكنة. أي أن الوسط والتباين والتغاير الخاص بـ u_t لا تتغير بمرور الزمن. إذا كانت |p| أقل من الواحد، فإنه يتضح من (4.2.12) أن قيمة التغاير ستقل. كلما اقتربنا في الفترة الزمنية سنرى أهمية تلك الخاصية في النتائج التالية بعد قليل.

أحد أسباب استخدام AR(1) ليس فقط سهولته مقارنةً بالنماذج الأعلى في الرتبة، ولكن أيضًا لأن نموذج AR(1) أثبت فعاليته في العديد من الأمثلة التطبيقية. بالإضافة إلى أن هناك العديد من الدراسات النظرية والعملية التي اهتمت بنموذج AR(1).

والآن بالعودة إلى نموذجنا الخاص بمتغيرين اثنين فقط: $Y_i = \beta_1 + \beta_2 X_t + u_t$ في الفصل 3 رأينا أن مقدر OLS لمعامل الميل هو:

$$\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum x_t^2} \qquad \textbf{(6.2.12)}$$

وتباينه معطى كالتالي:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2} \qquad \textbf{(7.2.12)}$$

حيث إن الحروف الصغيرة تعبر عن الانحرافات عن القيم الوسيطة.

والآن وفقًا للعملية AR(1)، يمكن إثبات أن تباين هذا المقدر يمكن كتابته كالتالي:

$$\text{var}(\hat{\beta}_2)_{\text{AR1}} = \frac{\sigma^2}{\sum x_t^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \cdots + 2\rho^{n-1}\frac{x_1 x_n}{\sum x_t^2}\right] \qquad \textbf{(8.2.12)}$$

حيث $\text{var}(\hat{\beta}_2)_{\text{AR1}}$ يعني تباين $\hat{\beta}_2$ وفقًا لنموذج الانحدار الذاتي من الدرجة الأولى.

بمقارنة (8.2.12) مع (7.2.12) نجد أن السابق يساوي التالي مضروبًا في مقدار يعتمد على ρ ، بالإضافة إلى الارتباط الذاتي في العينة بين قيم المتغير المنحدر X عند فترات زمنية مختلفة.(10)

وبوجه عام، لا تستطيع تحديد ما إذا كان $\text{var}(\hat{\beta}_2)$ أقل أو أكبر من $\text{var}(\hat{\beta}_2)_{\text{AR1}}$ [ولكن انظر المعادلة (1.4.12) التالية]. بالطبع إذا كانت ρ تساوي الصفر، فإن المعادلتين تتطابقان، كما يجب أن يحدث (لماذا؟). أيضًا، إذا كان الارتباط بين القيم المتتالية للمتغير المنحدر صغيراً جداً، فإن تباين OLS التقليدي لمقدر الميل لن يكون متحيزاً بشكل كبير. ولكن كقاعدة عامة، التباينان الاثنان لن يكونا متساويين.

لإلقاء مزيد من الضوء على الفرق بين التباين المعطى في (7.2.12) و (8.2.12)، افترض أن المتغير المنحدر X يتبع أيضًا عملية ارتباط ذاتي من الدرجة الأولى بمعامل ارتباط ذاتي r. إذن من الممكن إثبات أن (8.2.12) يمكن كتابتها كالتالي:

$$\text{var}(\hat{\beta}_2)_{\text{AR(1)}} = \frac{\sigma^2}{\sum x_t^2}\left(\frac{1+r\rho}{1-r\rho}\right) = \text{var}(\hat{\beta}_2)_{\text{OLS}}\left(\frac{1+r\rho}{1-r\rho}\right) \qquad \textbf{(9.2.12)}$$

فمثلاً إذا كان $r = 0.6$ و $\rho = 0.8$ فإن باستخدام (9.2.12) غير أن $\text{var}(\hat{\beta}_2)_{\text{AR1}} = 2.8461\ \text{var}(\hat{\beta}_2)_{\text{OLS}}$.

وبعبارة أخرى فإن $\text{var}(\hat{\beta}_2)_{\text{OLS}} = \frac{1}{2.8461}\ \text{var}(\hat{\beta}_2)_{\text{AR1}} = 0.3513\ \text{var}(\hat{\beta}_2)_{\text{AR1}}$

أي أن المعادلة التقليدية لـ OLS [أي (7.2.12)] ستقدر تباين $(\hat{\beta}_2)_{\text{AR1}}$ بأقل من قيمته بحوالي 65%. وكما سنرى الإجابة محددة وفقًا للقيم المعطاة لكل من r وρ. وأهمية هذا التطبيق هو تحذير القارئ من الاستخدام الأعمى لـ OLS التقليدية لحساب التباين والأخطاء المعيارية لمقدرات OLS يمكن أن يؤدي إلى نتائج خاطئة تمامًا.

افترض أننا مازلنا نستخدم مقدر OLS للـ $\hat{\beta}_2$، وتم تعديل معادلة التباين لنضع في الاعتبار نموذج AR(1). أي أننا نستخدم المعطاة في (6.2.12) ولكن بالتباين المعطى في (8.2.12). ما هي الآن الخصائص المتعلقة بـ $\hat{\beta}_2$ ؟ من السهل إثبات أن $\hat{\beta}_2$ مازال خطيًا وغير متحيز. وفى واقع الأمر، كما سنرى في الملحق A3 الفقرة 2.A3 فإن الفروض الخاصة بعدم وجود ارتباط تسلسلي، وفرض ثبات التباين غير مطلوبين ولا

(10) لاحظ أن المقدار $r = \sum x_t x_{t+1} / \sum x_t^2$ يمثل الارتباط بين X_t و X_{t+1} (أو X_{t-1}، حيث إن الارتباط متماثل)، $r_2 = \sum x_t x_{t+2} / \sum x_t^2$ يمثل الارتباط بين قيم الـ X's على فترتين زمنيتين وهكذا.

نحتاج إليهما لإثبات أن $\hat{\beta}_2$ مقدر غير متحيز. هل $\hat{\beta}_2$ مازال BLUE؟ للأسف لا ليس BLUE. ففي فئة المقدرات الخطية غير المتحيزة، ليس لديه أقل تباين. باختصار، $\hat{\beta}_2$، على الرغم من أنه مقدار خطي غير متحيز، إذ إنه غير كفء (بشكل نسبي بالطبع). سيجد القارئ أن هذه النتيجة قريبة من انخفاض كفاءة $\hat{\beta}_2$ في حالة وجود اختلاف في التباين. وهنا نجد أهمية لمقدرات المربعات الصغرى المرجحة لـ $\hat{\beta}_2^*$ المعطاة في (8.3.11) والتي تعتبر حالة خاصة من مقدرات المربعات الصغرى العامة (GLS) والتي تعتبر مقدرات كفء. في حالة وجود ارتباط ذاتي، هل ممكن أن نجد مقدر BLUE؟ الإجابة بنعم كما سنرى من المناقشة الآتية في الفقرة التالية.

3.12 المقدر BLUE في حالة وجود ارتباط ذاتي: THE BLUE ESTIMATOR IN THE PRESENCE OF AUTOCORRELATION

بالعودة إلى النموذج ثنائي المتغيرات، وبافتراض العملية AR(1)، يمكن أن نثبت أن المقدر الـ BLUE لـ β_2 معطى في المعادلة التالية:[11]

$$\hat{\beta}_2^{\text{GLS}} = \frac{\sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + C \qquad (1.3.12)$$

حيث C عامل الارتباط والذي يمكن إهماله عند التطبيق. لاحظ أن الترميز t يبدأ الآن من $t = 2$ إلى $t = n$. والتباين معطى كالتالي:

$$\text{var}\,\hat{\beta}_2^{\text{GLS}} = \frac{\sigma^2}{\sum_{t=2}^{n}(x_t - \rho x_{t-1})^2} + D \qquad (2.3.12)$$

حيث إن D عامل ارتباط أيضًا يمكن إهماله عند التطبيق (انظر تمرين 18.12) المقدر $\hat{\beta}_2^{\text{GLS}}$، كما يقترح اسمه تم الحصول عليه من طريقه GLS. كما وضحنا في الفصل 11، في GLS تتم إضافة أي معلومة جديدة تم الحصول عليها (مثلاً: عدم ثبات التباين أو الارتباط الذاتي) مباشرة في عملية التقدير عن طريق تحويل المتغيرات، في حين أنه في الـ OLS مثل هذه المعلومات لا يتم اعتيادها مباشرًا في عملية التقدير. وكما يرى القارئ فإن مقدر GLS الخاص بـ β_2 والمعطى في (1.3.12) يأخذ في اعتباره معامل الارتباط الذاتي ρ في عملية التقدير، أما معادلة OLS المعطاة

(11) للإثبات انظر Jan Kmenta, Elements of Econometrics, Macmillan, New York, 1971, pp. 274–275. معامل الارتباط C خاص بالمشاهدة الأولى (Y_1، X_1). لهذه النقطة انظر تمرين 18.12.

في (6.2.12) لا تعتبر ذلك وتتجاهله تمامًا. وهذا هو السبب الذي يجعل مقدر GLS يعتبر مقدر BLUE ، على عكس مقدر الـ OLS، حيث إن مقدر GLS يعتمد على الاستخدام الأمثل لكل المعلومات المتاحة عن المتغيرات.[12] ونلاحظ أنه إذا كانت $\rho = 0$ أي لا توجد أي معلومة إضافية يمكن اعتبارها، فإن مقدرات OLS و GLS ستتطابق تمامًا.

وبالتالي فباختصار، مقدر GLS المعطى في (1.3.12) هو المقدر الـBLUE وأقل التباين معطى في (2.3.12) وليس (8.2.12) وبالطبع ليس أيضًا (7.2.12).

ملاحظة فنية. كما لاحظنا في الفصل السابق، نظرية Gauss-Markov تعطي فقط الشروط الكافية للـ OLS حتى تصبح BLUE. الشروط الضرورية والكافية أيضًا لـ OLS حتى تصبح BLUE معطاة أيضًا في نظرية Krushkal والمذكورة في الفصل السابق. وبالتالي في بعض الحالات تكون OLS مقدرات BLUE حتى إذا كان هناك ارتباط ذاتي، ولكن هذه الحالات نادرة وغير متكررة كثيرًا عند التطبيق العملي.

ما الذي سيحدث إذا استخدمنا طريقة OLS التقليدية حتى في حالة وجود ارتباط ذاتي؟ الإجابة معطاة في الفقرة التالية.

4.12 عواقب استخدام OLS في حالة وجود ارتباط ذاتي:
CONSEQUENCES OF USING OLS IN THE PRESENCE OF AUTOCORRELATION

في حالة وجود ارتباط ذاتي، فإن مقدرات OLS خطية وغير متحيزة، كما كان الحال في وجود اختلاف في التباين، وتكون أيضًا متسقة وتؤول تقاربيًا إلى التوزيع الطبيعي. ولكن لم تعد مقدرات كفء (أي ليس لها أقل تباين). ما الذي يحدث في عملية اختبارات الفروض إذا استخدمنا مقدرات OLS في حالة وجود ارتباط ذاتي؟ مرة أخرى - كم كان الحال في وجود اختلاف في التباين - فإننا نفرق بين حالتين اثنتين. سنظل نستخدم نموذج الانحدار الذي يشتمل على متغيرين اثنين فقط.

(12) الإثبات النظري التقليدي لأن $\hat{\beta}_2^{GLS}$ مقدر BLUE موجود في ibid و Kmenta ولكن الخطوات الجبرية المعقدة ممكن تبسيطها بالاعتماد على نظام المصفوفات. انظر J. Johnston, Econometric Methods, 3d ed., McGraw-Hill, New York, 1984, pp. 291–293.

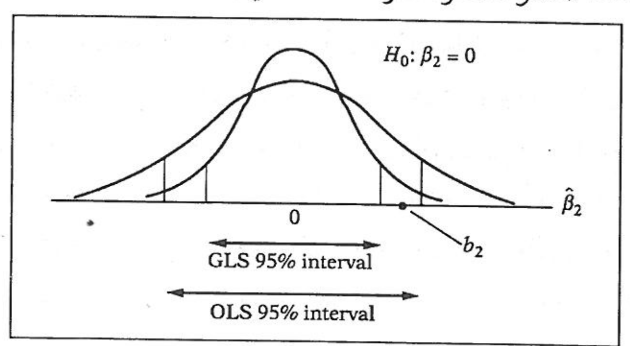
للتبسيط وتعميم المثال على نموذج انحدار متعدد به أكثر من متغيرين يعتبر أمراً سهلاً، ولا توجد فيه أي تعقيدات.[13]

تقدير OLS مع وجود ارتباط ذاتي: OLS Estimation Allowing for autocorrelation

كما سبق وذكرنا، فإن $\hat{\beta}_2$ ليست مقدر BLUE، حتى إذا استخدمنا $\text{var}(\hat{\beta}_2)_{\text{AR1}}$، ففترات الثقة المشتقة من ذلك ستكون أوسع من تلك المشتقة من طريقة GLS. وكما أثبت Kmenta، هذه النتيجة تظل كما هي حتى مع زيادة حجم القيمة بشكل ملحوظ.[14]

أي أن $\hat{\beta}_2$ ليس كفئاً بشكل تقاربي. نتائج هذه الإثباتات على اختبارات الفروض واضحة: فإننا سنتجه إلى إقرار أن المعامل غير معنوي إحصائيًا (أي لا يختلف عن الصفر) حتى وإن كان ذلك غير متحقق في الواقع (أي وفقًا لتقديرات GLS). الفرق ممكن ملاحظته بسهولة في الشكل (4.12). في هذا الشكل يوضح فترات الثقة وفقًا لـ 95% درجة ثقة لكل من OLS و GLS بافتراض أن قيمة β_2 الحقيقية تساوي الصفر. اعتبر تقدير ما لـ β_2 وهو مثلاً b_2 يقع في فترة OLS، سنقبل الفرض الخاص بأن β_2 يساوي الصفر بدرجة ثقة 95%. ولكن إذا استخدمنا فترة الثقة الخاصة بـ GLS (الصحيحة) سنرفض الفرض العدمي الخاص بأن قيمة β_2 الحقيقية تساوي الصفر، حيث إن b_2 تقع في منطقة الرفض.

إذن الخلاصة هي: لتكون فترات ثقة وتختبر بعض الفروض الإحصائية، لابد من استخدام GLS وليس OLS حتى مع العلم بأن مقدرات OLS هي مقدرات غير متحيزة ومتسقة. (عمومًا انظر الفقرة 11.12 التالية).



شكل (4.12) 95% فترات ثقة وفقًا لـ GLS و OLS : **GLS and OLS 95% Confidence Intervals**

(13) ولكن استخدام جبر المصفوفات يصبح ضرورة لذلك حتى نتفادى به تعقيدات الكثير من المعادلات الجبرية.

(14) انظر Kmenta, op. cit., pp. 277–278.

تقديرات OLS بغض النظر عن الارتباط الذاتي:
OLS Estimation Disregarding Autocorrelation

سيكون الموقف شديد الخطورة خصوصًا إذا قمت ليس فقط باستخدام $\hat{\beta}_2$ ولكن استخدمت $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$ والذي يتجاهل تمامًا مشكلة الارتباط الذاتي، ونكون هنا متصورين خطأ أن فروض النموذج التقليدية متحققة. **ستظهر الأخطاء وفقًا للأسباب التالية:**

1 - تباين البواقي $\hat{\sigma}^2 = \sum \hat{u}_t^2/(n-2)$ غالبًا ما سيقدر σ^2 بأقل من قيمتها الحقيقية.

2 - وكنتيجة لذلك، ستكون قيمة R^2 أعلى من قيمتها الحقيقية.

3 - حتى إذا لم تكن σ^2 مقدرة بأقل من قيمتها. فإن $\text{var}(\hat{\beta}_2)$ سيقدر $\text{var}(\hat{\beta}_2)_{\text{AR1}}$ بأقل من قيمته [المعادلة (8.2.12)]، وسيكون هذا هو التباين وفقًا للارتباط الذاتي من الدرجة الأولى حتى ولو كان الأخير غير كفء مقارنةً بـ $\text{var}(\hat{\beta}_2)^{\text{GLS}}$.

4 - وبالتالي قيم اختبارات المعنوية لـ t و F لم تعد صالحة للاستخدام، وإذا تم استخدامها ستعطي نتائج غير صحيحة ومضللة خاصة بالمعنوية الإحصائية لمعاملات الانحدار المقدرة.

لتوضيح السابق، دعنا نستخدم نموذج الانحدار ذا المتغيرين. ونحن نعلم من (الفصل 3) أنه وفقًا للفروض التقليدية فإن

$$\hat{\sigma}^2 = \frac{\sum \hat{u}_i^2}{(n-2)}$$

وفقًا للمقدر غير المتحيز للـ σ^2، أي أن $E(\hat{\sigma}^2) = \sigma^2$ ولكن إذا كان هناك ارتباط ذاتي، مثلاً AR(1)، سنرى أن:

$$E(\hat{\sigma}^2) = \frac{\sigma^2\{n - [2/(1-\rho)] - 2\rho r\}}{n-2} \tag{1.4.12}$$

حيث $r = \frac{\sum_{t=1}^{n-1} x_t x_{t-1}}{\sum_{t=1}^{n} x_t^2}$ والذي يمكن تفسيره كمعامل ارتباط (العينة) بين القيم المتالية للـ X's.[15] إذا كان كل من p و r موجبتين (ليس فرضًا منطقيًا في العديد من السلاسل الزمنية الاقتصادية) يظهر من (1.4.12) أن $E(\hat{\sigma}^2)<\sigma^2$ أي أن تباين البواقي

---

(15) انظر S. M. Goldfeld and R. E. Quandt, Nonlinear Methods in Econometrics, North Holland Publishing Company, Amsterdam, 1972, p. 183.

ودعنا نلاحظ بشكل عابر أنه إذا كانت الأخطاء مرتبطة طرديًا فإن قيمة R^2 ستكون أزيد من قيمتها الفعلية (متغيرة بشكل مرتفع) أي أنها تكون أكبر من قيمتها في حالة عدم وجود مثل هذا الارتباط.

المعتاد في المتوسط سيكون أقل من قيمة σ^2 الحقيقية. بمعنى آخر فإن $\hat{\sigma}^2$ سيكون متحيزاً بشكل منخفض. ولا نحتاج إلى أن نقول بأن هذا التحيز في $\hat{\sigma}^2$ سيظهر أيضًا في $\text{var}(\hat{\beta}_2)$ حيث إننا في الواقع نقدر الأخير باستخدام المعادلة .

ولكن حتى إذا لم يكن تقدير σ^2 بأقل من قيمته، فإن $\text{var}(\hat{\beta}_2)$ هو مقدر متحيز لـ $\text{var}(\hat{\beta}_2)_{AR1}$، ويمكن ملاحظة ذلك مباشرة من مقارنة (7.2.12) و (8.2.12). [16] حيث إن المعادلتين مختلفتان.

وفي حقيقة الأمر، إذا كانت ρ موجبة (وذلك يتحقق في العديد من السلاسل الزمنية الاقتصادية) وX's مرتبطة طرديًا (وذلك أيضًا يتحقق في معظم السلاسل الزمنية الاقتصادية) فإن

$$\text{var}(\hat{\beta}_2) < \text{var}(\hat{\beta}_2)_{AR1} \tag{2.4.12}$$

أي أن تباين OLS التقليدي لـ $\hat{\beta}_2$ يقلل من تقدير التباين وفقًا لوجود AR(1). [انظر المعادلة (9.2.12)]. وبالتالي إذا استخدمنا $\text{var}(\hat{\beta}_2)$ سنقلل من جودة ودقة تقدير (أي تقلل من الأخطاء المعيارية) الخاصة بـ $\hat{\beta}_2$. كنتيجة لذلك، عند حساب نسبة t كالتالي $t = \hat{\beta}_2/\text{se}(\hat{\beta}_2)$ (وفقًا لاختبار فرض $\beta_2 = 0$)، سنقدر t بأكبر من قيمتها، وبالتالي المعنوية الإحصائية للتقدير β_2. هذا الموقف سيزداد سوءا إذا أضفنا أن σ^2 مقدرة بأقل من قيمتها العقلية كما سبق وأوضحنا.

لنرى كيف أن OLS غالبًا ما ستقدر σ^2 بأقل من قيمتها وكذلك تباين $\hat{\beta}_2$. دعنا نقوم بعمل تجربة Monte Corlo التالية. افترض أنه في النموذج ثنائي المتغيرات الخاص بنا كنا نعلم أن القيمة الحقيقية لـ $\beta_1 = 1$ و $\beta_2 = 0.8$. وبالتالي فإن PRF العشوائية هي:

$$Y_t = 1.0 + 0.8\,X_t + u_t \tag{3.4.12}$$

وبالتالي

$$E(Y_t \mid X_t) = 1 + 0.8\,X_t \tag{4.4.12}$$

والذي يمثل خط انحدار المجتمع الحقيقي. دعنا نفترض أن u_t يتبع عملية ارتباط ذاتي من الدرجة الأولى كالتالي:

$$u_t = 0.7\,u_{t-1} + \varepsilon_t \tag{5.4.12}$$

حيث ε_t مستوفاة لكل فروض OLS. نفترض أيضًا أن التقارير الخاصة بـ εt's تتبع

(16) للإثبات النظري انظر في Kmenta, op. cit., p. 281 .

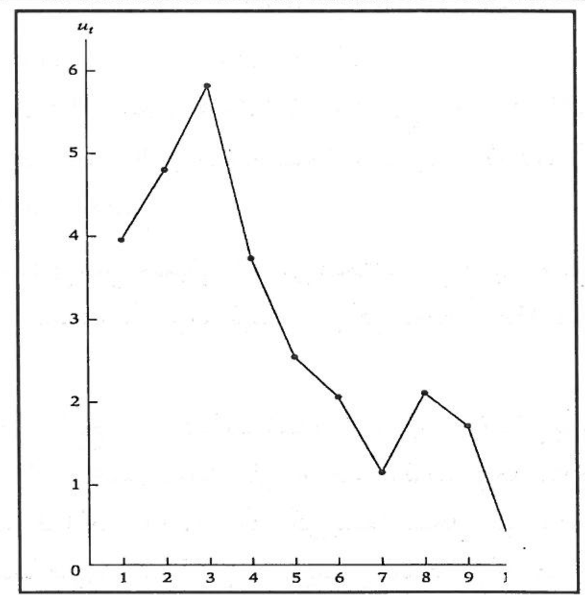
التوزيع الطبيعي بوسط يساوي الصفر وتباين الوحدة (=1). المعادلة (5.4.12) تعني أن الأخطاء المتتالية مرتبطة طرديًا، بمعامل ارتباط ذاتي يساوي 0.7+، وهو يمثل درجة عالية من الارتباط.

والآن باستخدام جدول توزيع طبيعي عشوائي بوسط = الصفر، وتباين يساوي الوحدة، دعنا نولد 10 أرقام عشوائية كالموضحة في الجدول (1.12) وباستخدام (5.4.12) نولد u_t. وحتى نبدأ هذه العملية نحتاج إلى نقطة مبدئية لـ u، مثلاً، $u_0 = 5$ ويرسم u_t المولدة في الجدول (1.12)، نحصل على الشكل (5.12) والذي يوضح أن كل قيمة تالية لـ u_t أكبر من سابقتها، وأصغر من القيمة التالية لها أي ارتباط ذاتي طردي.

جدول (1.12) : مثال افتراضي لحدود أخطاء مرتبطة طرديًا

A Hypothetical Example of positively Autocorrelated Error Terms

	ε_t^*	$u_t = 0.7u_{t-1} + \varepsilon_t$
0	0	$u_0 = 5$ (assumed)
1	0.464	$u_1 = 0.7(5) + 0.464 = 3.964$
2	2.026	$u_2 = 0.7(3.964) + 2.0262 = 4.8008$
3	2.455	$u_3 = 0.7(4.8010) + 2.455 = 5.8157$
4	−0.323	$u_4 = 0.7(5.8157) - 0.323 = 3.7480$
5	−0.068	$u_5 = 0.7(3.7480) - 0.068 = 2.5556$
6	0.296	$u_6 = 0.7(2.5556) + 0.296 = 2.0849$
7	−0.288	$u_7 = 0.7(2.0849) - 0.288 = 1.1714$
8	1.298	$u_8 = 0.7(1.1714) + 1.298 = 2.1180$
9	0.241	$u_9 = 0.7(2.1180) + 0.241 = 1.7236$
10	−0.957	$u_{10} = 0.7(1.7236) - 0.957 = 0.2495$



شكل (5.15) : ارتباط مولد وفقًا للعملية $u_t = 0.7\, u_{t-1} + \varepsilon_t$ [جدول (1.12)]

Correlation generated by the scheme $u_t = 0.7\, u_{t-1} + \varepsilon_t$ (Table 1.12)

جدول (2.12) : توليد قيم العينة لـ Y

X_t	u_t^*	$Y_t = 1.0 + 0.8X_t + u_t$
1	3.9640	$Y_1 = 1.0 + 0.8(1) + 3.9640 = 5.7640$
2	4.8010	$Y_2 = 1.0 + 0.8(2) + 4.8008 = 7.4008$
3	5.8157	$Y_3 = 1.0 + 0.8(3) + 5.8157 = 9.2157$
4	3.7480	$Y_4 = 1.0 + 0.8(4) + 3.7480 = 7.9480$
5	2.5556	$Y_5 = 1.0 + 0.8(5) + 2.5556 = 7.5556$
6	2.0849	$Y_6 = 1.0 + 0.8(6) + 2.0849 = 7.8849$
7	1.1714	$Y_7 = 1.0 + 0.8(7) + 1.1714 = 7.7714$
8	2.1180	$Y_8 = 1.0 + 0.8(8) + 2.1180 = 9.5180$
9	1.7236	$Y_9 = 1.0 + 0.8(9) + 1.7236 = 9.9236$
10	0.2495	$Y_{10} = 1.0 + 0.8(10) + 0.2495 = 9.2495$

* تم الحصول عليه من جدول (1.12)

الآن دعنا نفترض أن قيم الـ X مثبتة عند 1، 2، 3، . . . ، 10.

وبالتالي، وفقًا لهذه القيم، يمكن أن نولد قيم Y للقيمة من (3.4.12) وقيم u_t المعطاة في جدول (1.12). التفاصيل معطاة في جدول (2.12). باستخدام بيانات جدول (2.12)، إذا قمنا بعمل انحدار لـ Y على X، سنحصل على الانحدار (العينة) التالي:

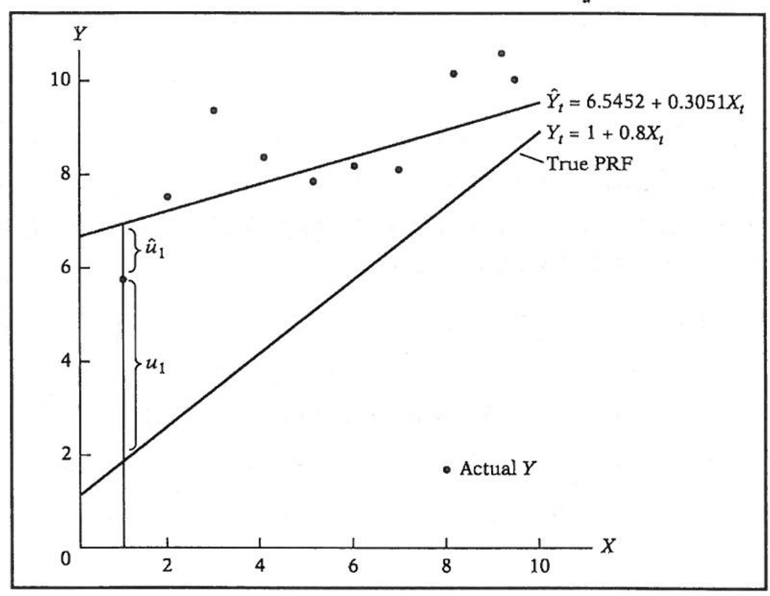
$$\hat{Y}_t = 6.5452 + 0.3051X_t$$
$$(0.6153) \quad (0.0992)$$
$$t = (10.6366) \quad (3.0763) \qquad (6.4.12)$$
$$r^2 = 0.5419 \qquad \hat{\sigma}^2 = 0.8114$$

في حين أن معادلة الانحدار الحقيقية معطاة في (4.4.12). الانحداران الاثنان ممثلان في الشكل (6.12)، والذي يوضح بشكل كبير قيم خطأ الانحدار المقدر بعيدة تمامًا عن خط الانحدار الحقيقي.

فتم تقدير معامل الميل الحقيقي بأقل من قيمته، وتم تقدير معامل الجزء الثابت المقطوع من المحور الصادي بأكبر من قيمته. (ولكن لاحظ أن مقدرات OLS مازالت غير متحيزة).

شكل (6.12) يوضح أيضًا لماذا يتم تقدير تباين u_i الحقيقي بأقل من قيمته باستخدام المقدر $\hat{\sigma}^2$، والمحسوب بناءً على $\hat{u}_i$. قيم $\hat{u}_i$ قريبة بوجه عام من الخط المقدر للانحدار (وذلك وفقًا لعملية الـ OLS) ولكن يبتعد بشكل كبير عن PRF الحقيقي. وبالتالي ليس لدينا صورة سليمة عن u_i. لمزيد من التفاصيل الخاصة بتقدير σ^2 بأقل من قيمتها الحقيقية، افترض أننا قمنا بتجربة أخرى وبالاحتفاظ بقيم X_t و ε_t من

الجداول (1.12) و (2.12)، دعنا نفترض أن $\rho = 0$، أي لا يوجد ارتبط ذاتي. قيم Y الجديدة المولدة معطاة في جدول (3.12).



شكل (6.12) : PRF الحقيقية وخط الانحدار المقدر لبيانات جدول (2.12)

جدول (3.12) : عينة من قيم الـ Y في حالة وجود صفر ارتباط تسلسلي

True PRF and the estimated Regression line for the data of table (12.2)

X_t	$\varepsilon_t = u_t^*$	$Y_t = 1.0 + 0.8X_t + \varepsilon_t$
1	0.464	2.264
2	2.026	4.626
3	2.455	5.855
4	−0.323	3.877
5	−0.068	4.932
6	0.296	6.096
7	−0.288	6.312
8	1.298	8.698
9	0.241	8.441
10	−0.957	8.043

* حيث إنه لا يوجد ارتباط ذاتي ، فإن قيم u_t و ε_t متطابقة قيم الـ ε_t من جدول (1.12) .

الانحدار وفقًا لجدول (3.12) يكون كالتالي :

$$\begin{aligned}\hat{Y}_t &= 2.5345 + 0.6145X_t \\ &\quad (0.6796) \quad (0.1087) \\ t &= (3.7910) \quad (5.6541) \\ &\qquad r^2 = 0.7997 \qquad \hat{\sigma}^2 = 0.9752\end{aligned} \tag{7.4.12}$$

هذا الانحـدار هو الأقـرب من الانحـدار «الحـقـيـقي»، حـيث إن قـيم الـ Y's عـشـوائيـة . لاحظ أن $\hat{\sigma}^2$ زادت عن 0.8114 $(\rho = 0.7)$ إلى 0.9752 $(\rho = 0)$. لاحظ أن الأخطاء المعيارية لـ $\hat{\beta}_1$ و$\hat{\beta}_2$ زادت أيضًا. هذه النتائج العملية متفقة مع النتائج النظرية التي تم استعراضها سابقًا.

5.12 العلاقة بين الأجور والإنتاجية في قطاع الأعمال بالولايات المتحدة الأمريكية، 1959-1998: RELATIONSHIP BETWEEN WAGES AND PRODUCTIVITY IN THE BUSINESS SECTOR OF THE UNITED STATES, 1959-1998

والآن بعد أن استعرضنا عواقب وجود ارتباط ذاتي. السؤال الأهم هو كيف يمكننا معرفة ما إذا كان هناك ارتباط ذاتي في البيانات أم لا؟ وكيف يمكن تعديل الأمر؟ قبل استعراض ذلك دعنا نأخذ هذا المثال التطبيقي. جدول (4.12) يعطي بيانات عن مؤشرات التعويض الحقيقي بالساعة (Y) والناتج بالساعة (X) في قطاع الأعمال في الاقتصاد الأمريكي خلال الفترة 1959 إلى 1998، أساس المؤشرات يبدأ 1992 = 100.

برسم بيانات Y على X، نحصل على الشكل (7.12). حيث إن العلاقة بين التعويضات الحقيقية وإنتاجية العمالة متوقع أن تكون علاقة موجبة (طردية)، فليس من المستغرب أن يكون هذان المتغيران مرتبطين طرديًا. المستغرب هو أن العلاقة بين المتغيرين تقريبًا خطية، على الرغم من وجود التصورات بأنه عند القيم المرتفعة للإنتاجية تكون العلاقة بين المتغيرين غير خطية نسبيًا.

وبالتالي، لقد قررنا أن يكون التقدير لنموذج لوغاريتمي خطي وحصلنا على النتائج التالية:

$$\begin{aligned} \hat{Y}_t &= 29.5192 + 0.7136X_t \\ se &= (1.9423) \quad (0.0241) \\ t &= (15.1977) \quad (29.6066) \\ r^2 &= 0.9584 \quad d = 0.1229 \quad \hat{\sigma} = 2.6755 \end{aligned} \tag{1.5.12}$$

حيث d هو إحصاء Durbin-Watson وسنناقشه قريبًا.

$$\begin{aligned} \widehat{\ln Y_t} &= 1.5239 + 0.6716 \ln X_t \\ se &= (0.0762) \quad (0.0175) \\ t &= (19.9945) \quad (38.2892) \\ r^2 &= 0.9747 \quad d = 0.1542 \quad \hat{\sigma} = 0.0260 \end{aligned} \tag{2.5.12}$$

لأغراض تحليلية ، نطلق على كل من (1.5.12) و(1.5.12) انحدارات الأجور الإنتاجية .

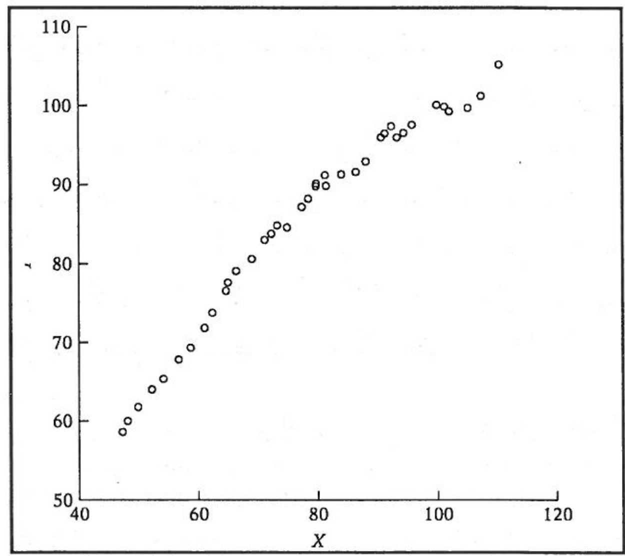
جدول (4.12) مؤشرات التعويضات الحقيقية والإنتاجية ، الولايات المتحدة ، 1959–1998

Observation	Y	X	Observation	Y	X
1959	58.5	47.2	1979	90.0	79.7
1960	59.9	48.0	1980	89.7	79.8
1961	61.7	49.8	1981	89.8	81.4
1962	63.9	52.1	1982	91.1	81.2
1963	65.3	54.1	1983	91.2	84.0
1964	67.8	54.6	1984	91.5	86.4
1965	69.3	58.6	1985	92.8	88.1
1966	71.8	61.0	1986	95.9	90.7
1967	73.7	62.3	1987	96.3	91.3
1968	76.5	64.5	1988	97.3	92.4
1969	77.6	64.8	1989	95.8	93.3
1970	79.0	66.2	1990	96.4	94.5
1971	80.5	68.8	1991	97.4	95.9
1972	82.9	71.0	1992	100.0	100.0
1973	84.7	73.1	1993	99.9	100.1
1974	83.7	72.2	1994	99.7	101.4
1975	84.5	74.8	1995	99.1	102.2
1976	87.0	77.2	1996	99.6	105.2
1977	88.1	78.4	1997	101.1	107.5
1978	89.7	79.5	1998	105.1	110.5

لاحظ أن : X = مؤشر الناتج في الساعة ، قطاع الأعمال (100=1992) .

Y = مؤشر التعويض الحقيقي في الساعة ، قطاع الأعمال (100=1992) .

المصدر : Economic Report of the President, 2000, Table B–47, p. 362 .



شكل (7.12) مؤشر التعويض (Y) ومؤشر الإنتاجية (X) ، الولايات المتحدة ، 1959–1998 .

Index of compensation (Y) and index of productivity (X), united states, 1959-1998

من الناحية النوعية، كل من النموذجين يعطيان نتائج متقاربة. في كل من الحالتين، المعاملات المقدرة لها معنوية عالية متمثلة في قيم t المرتفعة. في النموذج الخطي، إذا زاد مؤشر الإنتاجية بوحدة واحدة، فإنه في المتوسط، يرتفع مؤشر التعويض بحوالي 0.71 وحدة. في النموذج الخطي اللوغاريتمي، معامل الميل يعتبر معاملاً مرنًا (لماذا؟)، وجدنا أنه إذا زاد مؤشر الإنتاجية بـ 1%، فإنه في المتوسط يزداد مؤشر التعويض بحوالي 0.67%.

ما هي واقعية النتائج التي حصلنا عليها في (1.5.12) و (2.5.12) إذا كان هناك ارتباط ذاتي؟ كما سبق وذكرنا، إذا كان هناك ارتباط ذاتي، فإن الأخطاء المعيارية المقدرة تكون متحيزة، وكنتيجة لذلك رقم t المقدرة غير سليمة. وبالتالي نحن نحتاج إلى معرفة ما إذا كانت البيانات تعاني من الارتباط الذاتي أم لا. في الفقرة التالية سنناقش عددًا من الطرق التي تستخدم لاكتشاف الارتباط الذاتي، وسنشرح هذه الطرق بالنموذج الخطي (1.5.12) فقط، وسيترك النموذج اللوغاريتمي الخطي (2.5.12) كتمرين للقارئ.

6.12 اكتشاف الارتباط الذاتي: DETECTING AUTOCORRELATION

I - طريقة الرسم البياني: Graphical Method

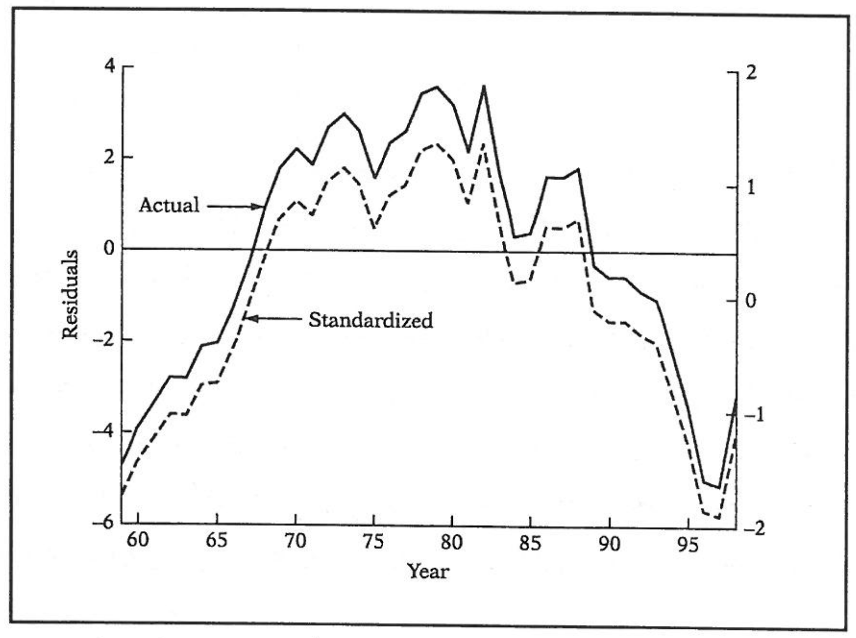
تذكر أن فرض عدم وجود الارتباط الذاتي في النموذج التقليدي ينطوي على أخطاء المجتمع u_t، والتي لا تشاهدها مباشرة. والذي يتوفر لدينا هو البواقي $\hat{u}_t$، والتي نحصل عليها من طريقة OLS التقليدية. وعلى الرغم من أن $\hat{u}_t$ ليست متطابقة مع u_t، [17] في الكثير من الأحيان، الاختبار النظري عن طريق الأشكال البيانية، للـ $\hat{u}$'s يعطينا تصورات أفضل عن وجود الارتباط الذاتي من عدمه. في الحقيقة، الاختبار النظري باستخدام الأشكال البيانية للـ $\hat{u}_t$ أو $(\hat{u}_i^2)$ يعطي معلومات مهمة ومفيدة ليس فقط للتحقق من وجود الارتباط الذاتي ولكن أيضًا اختلاف أو ثبات التباين

(17) حتى إذا كان u_t غير مرتبطة وثابتة التباين إلا أن البواقي $\hat{u}_t$ مرتبطة ذاتيًا وغير ثابتة التباين، للمزيد من التفاصيل، انظر G.S.Maddala, Introduction to Econometrics, 2d ed., Macmillan, New York, 1992, pp. 480–481,
عمومًا من الممكن إثبات أنه مع زيادة حجم العينة فإن البواقي تؤول إلى القيم الحقيقية للـ u_t's للمزيد من التفاصيل عن ذلك انظر E. Malinvaud, Statistical Methods of Econometrics, 2d ed., North-Holland Publishers, Amsterdam, 1970, p. 88.

(كما سنرى في الفصل القادم) بالإضافة إلى دقة النموذج وخطأ التوصيف، وسنستعرض كل ذلك في الفصل القادم.

أهمية عمل وتحليل الأشكال البيانية [للبواقي] كجزء أساسي في عملية التحليل الإحصائي، لا يجب أن تأخذ أكبر من حجمها. فعلى الرغم من أنها تعتبر وسيلة سهلة لفهم ملخص عن مشكلة معقدة، إلا أنها تعتبر اختبارًا لهذه البيانات كأنه اختبار تجميعي كلي، في حين أنه خاص فقط ببعض الحالات المنفردة. (18)

هناك العديد من الطرق لاختبار البواقي. ممكن أن نقوم برسم البواقي في مقابلة الزمن، ويسمى شكل التتابع الزمني، كما هو الحال في شكل (8.12)، والذي يوضح البواقي التي حصلنا عليها من انحدار الأجر - الإنتاجية (1.5.12)، قيم البواقي معطاة في جدول (5.12) بالإضافة إلى وجود بعض البيانات الأخرى.



شكل (8.12) البواقي والبواقي القياسية من انحدار الأجور - الإنتاجية (1.5.12)

Residuals and standardized residuals from the wages-productivity regression (12.5.1)

(18) Stanford Weisberg, Applied Linear Regression, John Wiley & Sons, New York, 1980, p. 120.

جدول (5.12) البواقي : الحقيقية ، القياسية والبواقي في فترات زمنية سابقة

Residuals: actual, standardized and lagged

Observation	RES1	SRES1	RES1(−1)	Observation	RES1	SRES1	RES1(−1)
1959	−4.703979	−1.758168		1979	3.602089	1.346324	3.444821
1960	−3.874907	−1.448293	−4.703979	1980	3.230723	1.207521	3.602089
1961	−3.359494	−1.255651	−3.874907	1981	2.188868	0.818116	3.230723
1962	−2.800911	−1.046874	−3.359494	1982	3.631600	1.357354	2.188868
1963	−2.828229	−1.057084	−2.800911	1983	1.733354	0.647862	3.631600
1964	−2.112378	−0.789526	−2.828229	1984	0.320571	0.119817	1.733354
1965	−2.039697	−0.762361	−2.112378	1985	0.407350	0.152252	0.320571
1966	−1.252480	−0.468129	−2.039697	1986	1.651836	0.617393	0.407350
1967	−0.280237	−0.104742	−1.252480	1987	1.623640	0.606855	1.651836
1968	0.949713	0.354966	−0.280237	1988	1.838615	0.687204	1.623640
1969	1.835615	0.686083	0.949713	1989	−0.303679	−0.113504	1.838615
1970	2.236492	0.835915	1.835615	1990	−0.560070	−0.209333	−0.303679
1971	1.880977	0.703038	2.236492	1991	−0.559193	−0.209005	−0.560070
1972	2.710926	1.013241	1.880977	1992	−0.885197	−0.330853	−0.559193
1973	3.012241	1.125861	2.710926	1993	−1.056563	−0.394903	−0.885197
1974	2.654535	0.992164	3.012241	1994	−2.184320	−0.816416	−1.056563
1975	1.599020	0.597653	2.654535	1995	−3.355248	−1.254064	−2.184320
1976	2.386238	0.891885	1.599020	1996	−4.996226	−1.867399	−3.355248
1977	2.629847	0.982936	2.386238	1997	−5.137643	−1.920255	−4.996226
1978	3.444821	1.287543	2.629847	1998	−3.278621	−1.225424	−5.137643

لاحظ أن: RES 1 = البواقي من انحدار (1.5.12)

$\frac{\text{RES 1}}{2.6755}$ = البواقي القياسية = SRES 1

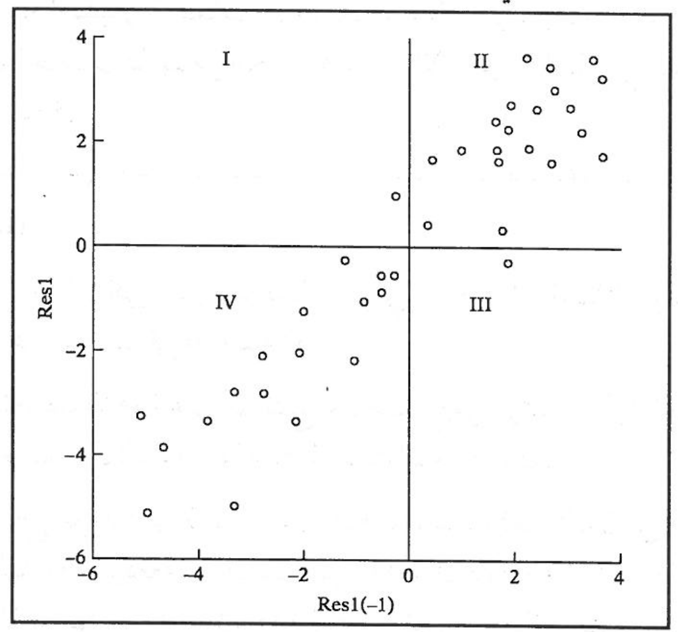
RES(−1) = البواقي في الفترة الزمنية السابقة الأولى .

كبديل يمكن أن تقوم برسم البواقي القياسية ضد الزمن، والموضحة في شكل (8.12) وجدول (5.12). البواقي القياسية هي ببساطة البواقي ($\hat{u}_t$) مقسومة على الأخطاء القياسية للانحدار ($\sqrt{\hat{\sigma}^2}$)، أي أنها تساوي ($\hat{u}_t / \hat{\sigma}$). لاحظ أن $\hat{u}_t$ و $\hat{\sigma}$ مقامة بنفس وحدة القياس الخاص بالمتغير المنحدر عليه Y. قيم البواقي القياسية ستكون قيمًا مجردة (بغض النظر عن وحدة القياس) ويمكن مقارنتها مع البواقي القياسية لأي نماذج انحدار أخرى. بالإضافة إلى ذلك، فإن البواقي القياسية، مثل $\hat{u}_t$، لها توقع يساوي الصفر (لماذا؟) وتباين يساوي الوحدة تقريبا.[19] في العينات الكبيرة ($\hat{u}_t / \hat{\sigma}$) تؤول تقاربيًا إلى التوزيع الطبيعي بوسط يساوي الصفر وتباين الوحدة. في مثالنا الحالي، $\hat{\sigma} = 2.6755$.

(19) في واقع الأمر، يقال عنها بواقي Studentized لأن لها تباينًا يساوي الوحدة. ولكن في التطبيق العملي ستعطي البواقي القياسية نفس الصورة وبالتالي يمكن الاعتماد عليهم. للمزيد من التفاصيل، انظر Norman Draper and Harry Smith, Applied Regression Analysis, 3d ed., John Wiley & Sons, New York, 1998, pp. 207–208.

باختيار الرسم البياني الخاص بالتتابع الزمني في شكل (8.12)، نلاحظ أن كلاً من $\hat{u}_t$ و $\hat{u}_{t-1}$ القياسية يكون لديها نمط ما موضح في الشكل (d1.12) مما يعني أنه على الأرجح u_t ليست عشوائية.

بشكل آخر، يمكن أن نرسم $\hat{u}_t$ ضد $\hat{u}_{t-1}$ ، أي نرسم البواقي عند الزمن t ضد قيم البواقي نفسها عند الزمن (t-1)، وهذا اختبار عملي للعملية RA(1) إذا كانت البواقي غير عشوائية، لابد أن نحصل على صورة مماثلة للموجودة في شكل (3.12). هذا الشكل البياني الخاص بانحدار الأجور - الإنتاجية موضح في الشكل (9.12)، البيانات الخاصة به معطاة في جدول (5.12).



شكل (9.12) : البواقي الحالية ضد البواقي في فترة زمنية سابقة

Current residuals versus lagged Reslduals

كما يوضح الشكل البياني، غالبية البواقي متركزة في الربع الثاني (شمال شرق) والرابع (جنوب غرب) مما يعني أن هناك ارتباطًا طرديًا قويًا بين البواقي.

طريقة الرسم البياني السابق استعراضها على الرغم من قوتها وسهولتها، إلا أنها لها طبيعة نوعية، وتتأثر بالحكم الشخصي للباحث. ولكن هناك العديد من الاختبارات الكمية التي يمكن الاعتماد عليها كبديل للطريقة النوعية الخالصة. دعنا الآن نستعرض بعضًا من هذه الاختبارات.

II - اختبار الدفعات : The Runs Test

إذا اختبرنا الشكل (8.12) بدقة، سنجد أن هناك العديد من الخصائص، مبدئيًا لدينا العديد من البواقي السالبة، وبالتالي هناك سلسلة من البواقي الموجبة ثم العديد من البواقي السالبة. إذا كانت هذه البواقي عشوائية بشكل مطلق، هل كنا سنلاحظ هذا النمط؟ من الواضح أن الإجابة الصحيحة هي لا. هذه الطريقة هي أساس ما يسمى اختبار الدفعات، ويسمى أيضًا في بعض الأحيان اختبار Geary، وهو اختبار لا معالمي.[20]

لتوضيح اختبار الدفعات، دعنا نحدد أولاً أن الإشارات (+ أو -) الخاصة بالبواقي تم الحصول عليها من انحدار الأجور - الإنتاجية والمعطى في العمود الأول من جدول (5.12).

$$(---------)(+++++++++++++++++++++)(----------) \quad \mathbf{(1.6.12)}$$

أي أن لدينا 9 بواقي سالبة يتبعها 21 بواقي موجبة ثم تليها 10 بواقي سالبة، وكإجمالي عدد من البواقي 40 مشاهدة.

يمكن تعريف الدفعة على أنها تتابع غير مقطوع من الإشارات أو الاتجاهات مثل + و -. ونعرف طول الدفعة بأنها عدد المفردات الموجودة فيها.

في التتابع الموجود في (1.6.12)، هناك 3 دفعات، دفعة تتكون من 9 قيم سالبة (أي أن طولها = 9)، ودفعة من 21 قيمة موجبة (أي أن طولها = 21) ودفعة من 10 قيم سالبة (أي أن طولها = 10). لرؤية هذه الدفعات تم وضع الإشارات ما بين أقواس في (1.6.12) باختبار شكل الدفعات في تتابع تام العشوائية من المشاهدات، يمكن أن نشتق اختبار العشوائية للدفعات. ونسأل السؤال التالي: هل الـ3 دفعات المكونة من 40 مشاهدة في مثالنا التوضيحي الحالي تعتبر كثيرة أو قليلة مقارنةً بعدد الدفعات المتوقع في تتابع تام العشوائية مكون من 40 مشاهدة؟ إذا كان هناك العديد من

---

(20) في الاختبارات اللامعلمية لا يكون لدينا أي فروض عن التوزيع (الاحتمالي) للمشاهدات المسحوبة. لمزيد من التفاصيل عن اختبار Geary انظر R.C.Geary, "Relative Efficiency of Count Sign Changes for Assessing Residual Autoregression in Least Squares Regression," Biometrika, vol. 57, 1970, pp. 123–127.

الدفعات، فإن هذا يعني أن البواقي الموجودة في المثال تتغير إشارتها باستمرار، مما يعني أن هناك ارتباطًا تسلسليًا سالبًا [انظر شكل (a3.12)]. وبالمثل إذا كان هناك القليل من هذه الدفعات، فإن هذا يعني وجود ارتباط ذاتي موجب (طردي). كما هو الحال في الشكل (a3.12). بالإضافة إلى أن شكل (8.12) يقترح وجود ارتباط طردي بين البواقي.

الآن دع N = إجمالي المشاهدات = $N_1 + N_2$.

N_1 = عدد المشاهدات ذات الإشارة + (أي البواقي الموجبة).

N_2 = عدد المشاهدات ذات الإشارة − (أي البواقي السالبة).

R = عدد الدفعات.

وبالتالي تحت صحة الفرض العدمي القائل بأن المشاهدات المتتالية (أي البواقي) مستقلة، وبافتراض أن $N_1 > 10$ و $N_2 > 10$، فإن عدد الدفعات يؤول تقاربيًا إلى التوزيع الطبيعي بـ

$$\text{توقع Mean:} \quad E(R) = \frac{2N_1N_2}{N} + 1 \tag{2.6.12}$$

$$\text{تباين Variance:} \quad \sigma_R^2 = \frac{2N_1N_2(2N_1N_2 - N)}{(N)^2(N-1)}$$

لاحظ أن: $N = N_1 + N_2$

وفقًا للفرض العدمي وباتباع التوزيع الطبيعي، فإن لدينا

$$\text{Prob}\,[E(R) - 1.96\sigma_R \le R \le E(R) + 1.96\sigma_R] = 0.95 \tag{3.6.12}$$

أي أن هناك احتمال 95% أن الفترة السابقة ستحتوي على R. وبالتالي فإن لدينا القاعدة التالية:

قاعدة القرار: لا ترفض الفرض العدمي الخاص بالعشوائية بدرجة ثقة 95% إذا كانت R، عدد الدفعات، تقع داخل فترة الثقة السابقة، أرفض الفرض العدمي إذا كانت R المقدرةتقع خارج هذه الحدود (لاحظ أنه يمكنك أن تختار درجة الثقة التي ترغب فيها).

بالعودة إلى مثالنا الحالي، نفرض أن N_1، عدد القيم السابقة هو 19 و N_2، عدد القيم الموجبة، هو 21 و $R = 3$. باستخدام المعادلات المعطاة في (2.6.12) نحصل على

$$E(R) = 10.975$$
$$\sigma_R^2 = 9.6936 \qquad (4.6.12)$$
$$\sigma_R = 3.1134$$

95% فترة ثقة للـ R في مثالنا الحالي هي :

$$[\, 10.975 \pm 1.96\ (3.1143)\,] = (4.8728,\ 17.0722)$$

من الواضح أن هذه الفترة لا تحتوي على 3. وبالتالي نرفض الفرض العدمي الخاص بأن بواقي نموذج الأجور - الإنتاجية هي بواقي عشوائية وذلك بدرجة ثقة 95% .

بعبارات أخرى، هناك ارتباط ذاتي في البواقي .

كقاعدة عامة، إذا كان هناك ارتباط ذاتي طردي، فإن عدد الدفعات سيكون قليلاً، في حين أنه إذا كان هناك ارتباط ذاتي عكسي فعدد الدفعات سيكون كثيراً . بالطبع من (2.6.12) يمكن أن نحدد ما إذا كان لدينا أعداد دفعات قليلة أو أعداد دفعات كثيرة .

Swed و Eisenhart قاما بتكوين جداول خاصة بالقيم الحرجة للعدد المتوقع للدفعات في تتابع عشوائي من N مشاهدة إذا كانت N_1 أو N_2 أقل من 20 . هذه الجداول معطاة في الملحق D، جدول (6.D) باستخدام هذه الجداول، يمكن للقارئ أن يثبت أن البواقي في انحدار الأجور - الإنتاجية غير عشوائي، وبالأخص يوجد ارتباط طردي .

III - اختبار Durbin - Watson [21] :

أكثر الاختبارات شيوعًا لاكتشاف الارتباط الذاتي هو اختبار Durbin-Watson والذي قام بعمله كل من الإحصائيين Durbin و Watson ويعرف باسم إحصاء d لـ Durbin-Watson، ويعرف كالتالي :

$$d = \frac{\sum_{t=2}^{t=n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{t=n}\hat{u}_t^2} \qquad (5.6.12)$$

وهو ببساطة النسبة بين مجموع الفروق المربعة للبواقي المتتالية في الزمن إلى RSS . لاحظ أن في بسط الإحصاء d، عدد المشاهدات هو n-1 حيث إن هناك مشاهدة مفقودة عند التعامل مع الفروق المتتالية .

---

(21) J. Durbin and G. S. Watson, "Testing for Serial Correlation in Least-Squares Regression," Biometrika, vol. 38, 1951, pp. 159–171.

الفائدة الكبيرة وراء استخدام الإحصاء d ، تكمن في أنه مبني على أساس البواقي المقدرة، والتي تحسب بشكل تلقائي في تحليل الانحدار.

ونظراً لهذه الفائدة، فإنه من المعتاد الآن توثيق قيمة Durbin-Watson مع باقي الإحصاءات المهمة الأخرى، مثل R^2، R^2 المعدلة d، t و F وعلى الرغم من أنه يستخدم الآن بشكل كبير وتلقائي، إلا أنه من المهم التعرف على الفروض القائم عليها اختبار Durbin-Watson d :

1 - نموذج الانحدار يجب أن يحتوي على الجزء الثالث المقطوع من المحور الصادي إذا لم يكن موجوداً، كما هو الحال في الانحدار المار بنقطة الأصل، من المهم أن نعيد الانحدار مستخدمين جزءاً ثابتًا مقطوعا من المحور الصادي للحصول على RSS .[22]

2 - المتغيرات المفسرة، X's، غير عشوائية، أو ثابتة عند تكرار المعاينة.

3 - الخطأ u_t مولد عن طريق عملية انحدار ذاتي من الدرجة الأولى : $u_t = \rho u_{t-1} + \varepsilon_t$. وبالتالي لا يستطيع الاختبار اكتشاف عمليات انحدار ذاتي ذات رتب أعلى.

4 - مقدار الخطأ u_t مفترض أنه يتبع التوزيع الطبيعي.

5 - نموذج الانحدار لا يشتمل على قيم في فترات زمنية سابقة خاصة بالمتغير التابع كواحدة من قيم المتغيرات المفسرة. وبالتالي الاختبار لا يمكن تطبيقه في النماذج على الشكل التالي:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \dots + \beta_k X_{kt} + \gamma Y_{t-1} + u_t \qquad \textbf{(6.6.12)}$$

حيث Y_{t-1} هي نفس المتغير التابع Y ولكن في فترة زمنية سابقة واحدة.

هذه النماذج معروفة باسم نماذج الانحدار الذاتي، وسندرسها بالتفصيل في الفصل (17).

6 - يفترض هذا الاختبار عدم وجود مشاهدات مفقودة في البيانات، وبالتالي في نموذج الأجور - الإنتاجية للفترة 1959 إلى 1998 . إذا كان مثلاً هناك مشاهدات

(22) عموماً ، قام R. W. Farebrother بحساب رقم d في حالة عدم وجود الجزء الثابت المقطوع من المحور الصادي في النموذج . انظر في "The Durbin-Watson Test for Serial Correlation When There Is No Intercept in the Regression, "Econometrica, vol. 48, 1980, pp. 1553–1563.

مفقودة عند الفترة 1978 وأيضًا عند 1982 لأي سبب من الأسباب، فإن الإحصاء d لا يمكن تطبيقه لوجود مثل هذه المشاهدات المفقودة.[23]

التوزيع الاحتمالي للإحصاء d المعطى في (5.6.12) من الصعب الحصول عليه، حيث إن Durbin و Watson أثبتا أن ذلك يعتمد على طريقة معقدة خاصة بقيمة الـ X الموجودة في العينة.[24]

هذه الصعوبة من المتوقع أن تكون مفهومة، حيث إن d يحسب على أساس $\hat{u}_t$ والذي يعتمد بالطبع على X's المعطاة. وبالتالي على خلاف اختبارات t، F و χ^2، لا توجد قيم حرجة محددة تؤدي إلى قبول أو رفض الفرض العدمي والخاص بوجود ارتباط ذاتي من الدرجة الأولى للأخطاء u_i. عمومًا، نجح Durbin و Watson في الحصول على حد أدنى d_L وحد أعلى d_u بحيث إنهما يمثلان حدودًا للقيمة d المحسوبة من (5.6.12) بحيث إذا وقعت خارجهما يمكن أن يتخذ القارئ قرارًا محددًا بخصوص وجود أو عدم وجود ارتباط تسلسلي طردي أو عكسي. والأكثر من ذلك، فإن هذه الحدود تعتمد فقط على عدد المشاهدات n، وعدد المتغيرات المفسرة، ولا يعتمد على القيم التي تأخذها هذه المتغيرات المفسرة، هذه الحدود والخاصة بحجم العينة n من 6 إلى 200 وحتى 20 متغيرًا مفسرًا تم جدولتها بواسطة Durbin و Watson وموجود في الملحق D، جدول D.5 (حتى 20 متغيرًا مفسرًا فقط).

طريقة الاختبار الفعلية يمكن تفسيرها بشكل أكثر وضوحًا بالاعتماد على الشكل (10.12)، والذي يوضح هذه الحدود للـ d وهي 0 و 4. ويمكن توضيح ذلك كالتالي مع تطبيق (5.6.12) نحصل على

$$d = \frac{\sum \hat{u}_t^2 + \sum \hat{u}_{t-1}^2 - 2\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2} \qquad \textbf{(7.6.12)}$$

حيث $\Sigma\hat{u}_t^2$ و $\Sigma\hat{u}_{t-1}^2$ يختلف عن بعض بمشاهدة واحدة فقط وهما تقريبًا متساويان. وبالتالي يمكن استخدام $\Sigma\hat{u}_t^2 \approx \Sigma\hat{u}_{t-1}^2$ وبالتالي يمكن كتابة (7.6.12) كالتالي:

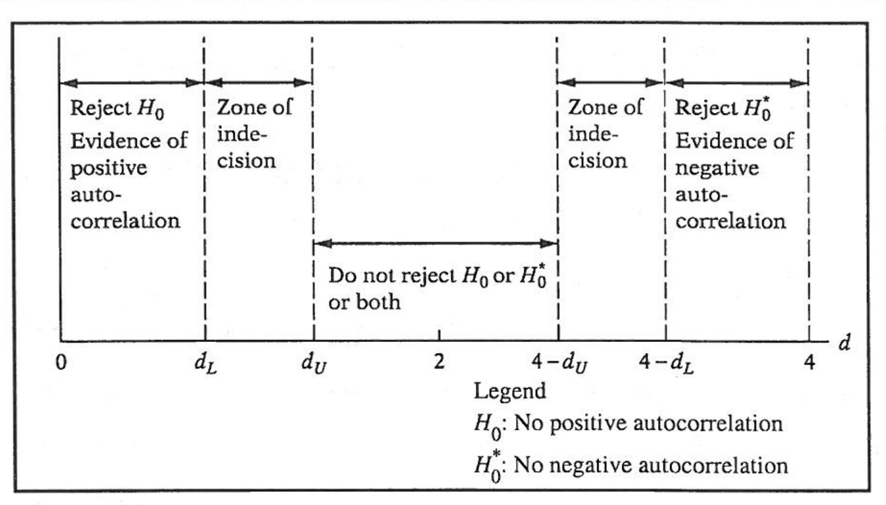
$$d \approx 2\left(1 - \frac{\sum \hat{u}_t\hat{u}_{t-1}}{\sum \hat{u}_t^2}\right) \qquad \textbf{(8.6.12)}$$

حيث $\approx$ تعني تقريبًا متساويان.

---

(23) لمزيد من التفاصيل، انظر Gabor Korosi, Laszlo Matyas, and Istvan P. Szekey, Practical Econometrics, Avebury Press. England, 1992, pp. 88–89.

(24) ولكن انظر في المناقشة الخاصة باختبار Durbin-Watson التام المعطى لاحقًا في هذه الفترة.



(شكل 10.12) إحصاء **Durbin-Watson *d***

والآن دعنا نستعرض التعريف التالي :

$$\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2} \qquad \textbf{(9.6.12)}$$

على أنها معامل الارتباط الذاتي للعينة من الدرجة الأولى، ونعتبره تقديراً لـ ρ (انظر هامش 9) باستخدام (9.6.12) يمكن أن نكتب (8.6.12) كالتالي :

$$d \approx 2(1-\hat{\rho}) \qquad \textbf{(10.6.12)}$$

ولكن بما أن $-1 \leq \rho \leq 1$ ، فإن (10.6.12) تعني أن :

$$0 \leq d \leq 4 \qquad \textbf{(11.6.12)}$$

هذه هي حدود الـ d، وأي قيمة مقدرة للـ d ستقع بين هذه الحدود .

يظهر من المعادلة (10.6.12) أنه إذا كانت $\hat{\rho}=0$ ، $d=2$، أي أنه لا يوجد ارتباط تسلسلي (من الدرجة الأولى)، فإن d متوقع أن تساوي تقريباً 2 . وبالتالي كقاعدة عامة إذا وجدت d تساوي 2 في أي تطبيق يمكن للباحث أن يفترض عدم وجود ارتباط ذاتي من الدرجة الأولى سواء طردي أو عكسي . أما إذا كانت $\hat{\rho} = +1$، مما يعني وجود ارتباط طردي تام في البواقي فإن $d \approx 0$. وبالتالي إذا كانت d قريبة من الصفر، فإن ذلك يمثل دليلاً أقوى على وجود ارتباط تسلسلي طردي . هذه العلاقة تظهر من (5.6.12) حيث إنه في حالة وجود ارتباط ذاتي طردي، فإن قيم $\hat{u}_t$'s ستزداد

معًا والفروق ستكون صغيرة، وكنتيجة لذلك، فإن البسط الذي يشتمل على مجموع مربعات الفروق سيكون صغيرًا مقارنةً مع المقام الذي يشتمل على مجموع المربعات والذي يظل ثابتًا وله قيمة محددة لكل تحليل انحدار.

إذا كانت $\hat{\rho} = -1$، أي أن هناك ارتباطًا ذاتيًا تامًا عكسيًا بين البواقي المتتالية، $d \approx 4$. وبالتالي كلما اقتربت d من 4 مما يعني أن هناك دليلاً أقوى على وجود ارتباط تسلسلي عكسي. مرة أخرى بالنظر إلى (5.6.12) نرى ذلك بشكل واضح.

إذا كان هناك ارتباط ذاتي عكسي، سنجد $\hat{u}_t$ الموجبة تتبعها $\hat{u}_t$ سالبة وبالعكس، وبالتالي سيكون $|\hat{u}_t - \hat{u}_{t-1}|$ دائمًا أكبر من $|\hat{u}_t|$. وبالتالي سيكون بسط الـ d أكبر عادة من المقام.

الطريقة التي يعمل بها اختبار Durbin-Watson يمكن توضيحها كالتالي مع افتراض تحقق كل الفروض الرئيسية لهذا الاختبار:

1 – قم بعمل انحدار OLS واحصل على البواقي.

2 – احسب d من (5.6.12) (معظم الحزم الإحصائية تقوم بحساب d تلقائيًا).

3 – وفقًا لحجم العينة وعدد المتغيرات المفسرة اوجد القيمة الحرجة لـ d_L و d_u.

4 – اتبع الآن قواعد القرار المعطاة في جدول (6.12).

لسهولة التطبيق، هذه القواعد الخاصة باتخاذ القرار، معطاة مرة أخرى في الشكل (10.12).

لشرح الطريقة، دعنا نعود إلى انحدار الأجور – الإنتاجية. من البيانات المعطاة في جدول (5.12) قيمة d المقدرة تساوي 0.1229، مما يعني أن هناك ارتباطًا تسلسليًا موجبًا (طرديًا) في البواقي. من جداول Derbin-Watson، نجد أن لدينا 40 مشاهدة ومتغيرًا مفسرًا واحدًا، وبالتالي $d_L = 1.44$ و $d_u = 1.54$ عند مستوى معنوية 5% بما أن d المحسوبة والمساوية لـ 0.1229 تقع أسفل d_L لن تستطيع رفض الفرض العدمي، أي أنه يوجد ارتباط تسلسلي طردي في البواقي.

على الرغم من الشهرة الكبيرة لاختبار d، إلا أن أهم عيوبه هو الوقوع في المنطقة التي لا تستطيع أن تأخذ فيها أي قرار، وبالتالي لا تستطيع استنتاج ما إذا كان الارتباط الذاتي (من الدرجة الأولى) موجودًا أو غير موجود.

لحل هذه المشكلة، العديد من الكتاب اقترحوا تعديلات على اختبار d ، ولكن نظرًا لتعقيدهم وبعدهم عن محتوى الكتاب الحالي، لن يتم استعراضهم في هذا الكتاب .[25] في العديد من المواقف، عمومًا وجد أن الحد الأعلى d_u هو الحد الحقيقي المعنوي، وبالتالي في حالة أن d تقع في منطقة اللاقرار، يمكن اتباع قاعدة اختبار d المعدل التالية : وفقًا لمستوى المعنوية α .

1 - $H_0 : \rho = 0$ ضد $H_1 : \rho > 0$. أرفض H_0 عند المستوى α إذا كان $d < d_u$. أي يوجد ارتباط ذاتي طردي له معنوية إحصائية .

جدول (6.12) اختبار Durbin - Watson d : قواعد القرار – **Durbin-Watson d Test: Decision Rules**

Null hypothesis	Decision	If
No positive autocorrelation	Reject	$0 < d < d_L$
No positive autocorrelation	No decision	$d_L \le d \le d_U$
No negative correlation	Reject	$4 - d_L < d < 4$
No negative correlation	No decision	$4 - d_U \le d \le 4 - d_L$
No autocorrelation, positive or negative	Do not reject	$d_U < d < 4 - d_U$

2 - $H_0 : \rho = 0$ ضد $H_1 : \rho < 0$ ، أرفض H_0 عند المستوى α إذا كان $(u - d) < d_u$ ، أي هناك معنوية إحصائية لوجود ارتباط ذاتي عكسي .

3 - $H_0 : \rho = 0$ ضد $H_1 : \rho \neq 0$ ، أرفض H_0 عند المستوى 2α إذا كان $d < d_u$ أو $(4 - d) < d_u$ ، أي هناك معنوية إحصائية لوجود ارتباط ذاتي طردي أمر عكسي .

من الواضح أيضًا أن المنطقة التي يصعب فيها اتخاذ القرار تقل مع زيادة حجم العينة، وذلك يتضح من جداول Durbin - Watson . فمثلاً في حالة وجود a متغيرات منحدرة و 20 مشاهدة، فإن 5% حد أدنى وأعلى للـ d هما 0.894 و 1.28 بالترتيب . ولكن هاتان القيمتين هما 1.515 و 1.739 إذا كان حجم القيمة 75 .

حزمة الحاسب الآلي Shazam تقوم بعمل الاختبار التام للـ d أي أنها تعطي قيمة p-value، أي الاحتمال الخاص بقيمة d المحسوبة . وبالتالي بالاعتماد على إمكانيات الحاسب الآلي، لم يعد صعبًا إيجاد قيمة p-value لقيمة d المحسوبة . باستخدام Shazam (النسخة 9) لانحدار الأجور الإنتاجية، نجد أن قيمة p-value لـ d المحسوبة هي 0.1229 أي قريبة من الصفر، مما يدعم استنتاجنا السابق بناء على جداول Durbin-Watson .

(25) لمزيد من التفاصيل، انظر Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, Advanced Econometric Methods, Springer Verlag, New York, 1984, pp. 225–228.

اختبار Durbin-Watson أصبح اختباراً شائع الاستخدام، لدرجة أن المستخدمين تنساوا الفروض التي يبنى عليها هذا الاختبار. بالتحديد هذه الفرض هي: (1) المتغيرات المفسرة أو المنحدرة غير عشوائية (2) حد الخطأ يتبع التوزيع الطبيعي و(3) نماذج الانحدار لا تشتمل على قيم للمتغير المنحدر في فترات زمنية سابقة، وهذا الفرض يعتبر فرضاً مهماً لتطبيق اختبار d.

إذا كان نموذج الانحدار يشتمل على قيم للمتغير المنحدرة عليه في فترات زمنية سابقة، فإن قيمة d في هذه الحالة تكون قريبة من 2، مما يعني عدم وجود ارتباط ذاتي (من الدرجة الأولى) في مثل هذه النماذج. وهذا يعني أن نماذج الارتباط الذاتي لا تعاني بالفعل من مشكلة الارتباط الذاتي. في واقع الأمر، فقد قام Durbin بعمل اختبار يسمى اختبار h لاختبار الارتباط التسلسلي في مثل هذه النماذج. ولكن هذا الاختبار ليس بنفس قوة، بمعنى إحصائي، اختبار Breusch-Godfrey الذي تم مناقشته سابقاً، وبالتالي ليس هناك حاجة لاستعراض اختبار h. ولكن نظراً لأهميته التاريخية فقد تمت مناقشة هذا الاختبار في تمرين 36.12.

بالإضافة إلى ذلك، إذا كان حد الخطأ u_t ليس AIID، فإن اختيار d قد لا يعطي نتائج سليمة.[26] في هذه الحالة، إن اختبار الدفعات السابق شرحه يكون من الأفضل استخدامه، حيث إنه لا يعتمد على أي فروض خاصة بالتوزيع الاحتمالي لحد الخطأ. عموماً إذا كان حجم العينة كبيراً (بشكل فني يمكن اعتباره غير محدود)، فمن الممكن استخدام Durbin-Watson d حيث يمكن إثبات أن[27]

$$\sqrt{n}\left(1-\frac{1}{2}d\right) \approx N(0,1) \qquad (12.6.12)$$

أي أنه في العينات الكبيرة وباستخدام التحويلة (12.6.12) وتطبيقها على إحصاء d فإنه سيتبع التوزيع الطبيعي القياسي. ونلاحظ أن العلاقة بين d و$\hat{\rho}$، معامل الارتباط الذاتي المقدر من الدرجة الأولى والموجودة في (10.6.12)، تتبع التالي:

$$\sqrt{n}\hat{\rho} \approx N(0,1) \qquad (13.6.12)$$

(26) لمزيد من التفاصيل الدقيقة، انظر Ron C. Mittelhammer, George G. Judge, and Douglas J. Miller, Econometric Foundations, Cambridge University Press, New York, 2000, p. 550.

(27) انظر James Davidson, Econometric Theory, Blackwell Publishers, New York, 2000, p. 161.

أي أنه في العينات الكبيرة، الجذر التربيعي لحجم العينة مضروب في معامل الارتباط الذاتي المقدر من الدرجة الأولى يتبع التوزيع الطبيعي القياسي.

لتوضيح الاختبار، وباستخدام مثال الأجور والإنتاجية وجدنا أن $d = 0.1229$ باستخدام $h = 40$. وبالتالي في (12.6.12) نجد أن

$$\sqrt{40}\left(1 - \frac{0.1229}{2}\right) \approx 5.94$$

اذا كان الفرض العدمي الخاص بالارتباط الذاتي من الدرجة الأولى يساوي الصفر صحيحًا ، فإن الاحتمال التقاربي لقيمة Z (المتغير الذي يتبع التوزيع الطبيعي القياسي) والذي يساوي 5.94 أو أكبر يكون صغيراً جداً. تذكر أنه في التوزيع الطبيعي القياسي 5% قيمة حرجة (اختبار ذي طرفين) تكون 1.96 و 1% قيمة حرجة للـ Z تكون 2.58. وعلى الرغم من أن حجم العينة يساوي 40، إلا أنه وفقًا لأغراض تطبيقية، فإنه قد يكون كبيرًا بشكل كاف لاستخدام التوزيع الطبيعي تقاربيًا. الاستنتاج يظل كما هو، وهو أن بواقي انحدار الأجور - الإنتاجية يعاني من مشكلة الارتباط الذاتي.

ولكن المشكلة الأكثر خطورة والخاصة باختبار d ، هو فرض أن المتغيرات المنحدرة غير عشوائية، أي أن قيمها ثابتة في العينات المتكررة. إذا لم يتحقق ذلك، فإن اختبار d لا يمكن استخدامه سواء كان حجم العينة صغيراً، محدوداً أو كبيراً وغير محدود.[28] وحيث إن هذا الفرض لا يمكن تحقيقه في الكثير من النماذج الاقتصادية التي تحتوي على بيانات سلاسل زمنية توصل أحد الكتاب إلى أن إحصاء Durbin- Watson قد لا يكون مفيداً في الاقتصاد القياسي الذي يحتوي على بيانات سلاسل زمنية.[29]

ووفقًا لذلك، فإن هناك اختبارات أكثر فائدة وأهمية في موضوع الارتباط الذاتي، ولكنها جميعًا تعتمد على العينات كبيرة الحجم، سنستعرض أحد هذه الاختبارات الآن وهو اختبار Breusch - Godfrey.

(28) Ibid., p. 161.

(29) Fumio Hayashi, Econometrics, Princeton University Press, Princeton, N. J., 2000, p. 45.

IV - اختبار عام للارتباط الذاتي: A General Test of Autocorrelation

اختبار Breusch-Godfrey (BG)[30] :

لتفادي بعض العيوب الموجودة في اختبار Durbin-Watson للارتباط الذاتي، قام كل من Breusch و Godfrey بعمل اختبار للارتباط الذاتي، والذي يعتبر اختبارا عامًا بمعنى أنه يسمح بـ (1) متغيرات منحدرة غير عشوائية مثل قيم للمتغير المنحدر عليه في فترات زمنية سابقة، (2) ارتباط ذاتي من درجة أعلى من الدرجة الأولى مثل AR(1)، AR(2) وهكذا، (3) نماذج المتوسطات المتحركة من الدرجة الأولى أو أعلى للحد الخطأ، مثل 3t الموجود في (1.2.13).[31]

وبدون الدخول في التفاصيل الرياضية، والتي يمكن الحصول عليها من المراجع، فإن اختبار BG والمعروف أيضًا باختبار LM،[32] يعمل كالتالي:

باستخدام نموذج انحدار ذي متغيرين اثنين لتوضيح الاختبار، على الرغم من أنه يمكن تطبيقه في نماذج الانحدار التي يشتمل على أكثر من متغيرين، ويمكن أيضًا إضافة المتغير المنحدر عليه في فترات زمنية سابقة. دعنا نفترض أن لدينا التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (14.6.12)$$

افترض أن حد الخطأ u_t يتبع نموذج انحدار ذاتي من الدرجة p، AR(p) كالتالي:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \dots + \rho_p u_{t-p} + \varepsilon_t \qquad (15.6.12)$$

حيث ε_t هي حد الخطأ العشوائي السابق ذكره. وكما نرى، فإن النموذج السابق ما هو إلا الصورة العامة لـ AR(1).

الفرض العدمي H_0 والمراد اختباره هو:

$$H_0 : \rho_1 = \rho_2 = \dots = \rho_p = 0 \qquad (16.6.12)$$

(30) انظر L. G. Godfrey, "Testing Against General Autoregressive nd Moving Average Error Models When the Regressor include Lagged Dependent variables," Econometrica, vol. 46, 1978, pp. 1293–1302, and T. S. Breusch, "Thesting for Autocorrelation in Dynamic Linear Models, Australian Economic Papers, vol. 17, 1978, pp. 334–355.

(31) فمثلاً، في الانحدار $Y_t = \beta_1 + \beta_2 X_t + u_t$، فإن حد الخطأ يمكن تمثيله كالتالي $u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2}$ والذي يمثل نموذج متوسطات متحركة من الدرجة الثانية لحد الخطأ ε_t .

(32) هذا الاختبار مبني على معامل Lagrange كما سبق وذكرنا في الفصل 8 .

أي أنه لا يوجد ارتباط تسلسلي من أي درجة . اختبار **BG** يعتمد على الخطوات التالية :

1 - قدر (14.6.12) باستخدام OLS واحصل على البواقي $\hat{u}_t$.

2 - قم بعمل انحدار لـ $\hat{u}_t$ على قيم X_t الأصلية (إذا كان هناك أكثر من X واحد قم بإدخالها جميعًا في النموذج) و $\hat{u}_{t-1}$ ، $\hat{u}_{t-2}$ ، . . . . ، $\hat{u}_{t-p}$ حيث إن ذلك يعتبر قيم البواقي المقدرة في الخطوة 1 في فترات زمنية سابقة ، وبالتالي إذا كانت $p = 4$ ، فإننا سيكون لدينا أربع قيم في فترات زمنية سابقة للبواقي كمتغيرات منحدرة في النموذج .

لاحظ أنه لكي نقوم بهذا الانحدار فإن لدينا $(n - p)$ مفردة (لماذا؟) .

للاختصار قم بعمل الانحدار التالي :

$$\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t \qquad \textbf{(17.6.12)}$$

واحصل على R^2 من هذا الانحدار (المساعد) .[33]

3 - إذا كان حجم القيمة كبيرًا (غير محدود) ، فإن Breusch و Godfrey أثبتا أن :

$$(n - p)R^2 \sim \chi^2_p \qquad \textbf{(18.6.12)}$$

أي أن $(n - p)$ مضروبة في قيمة R^2 التي نحصل عليها من الانحدار المساعد (17.6.12) تؤول تقاربيًا إلى توزيع كاى التربيعي بدرجة حرية p في أي تطبيق ، إذا زادت قيمة $(n - p)\ R^2$ عن القيمة الحرجة لكاى التربيعي عند مستوى المعنوية المحدد ، فإننا نرفض الفرض العدمي ، مما يعني أن لدينا على الأقل واحدة من الـ rho في (15.6.12) تختلف معنويًا عن الصفر .

والآن لدينا الملاحظات التالية والخاصة بتطبيق اختبار BG :

1 - المتغيرات المنحدرة الموجودة في نموذج الانحدار قد يكون منها قيم المتغير المنحدر عليه Y في فترات زمنية سابقة ، أي Y_{t-1} ، Y_{t-2} وهكذا ، ويكون موجودًا كمتغير

(33) السبب في استخدام المتغير المنحدر الأصلي X في النموذج هو السماح لكون X غير عشوائي بشكل تام . ولكن اذا كان غير عشوائي تمامًا فإنه من الممكن حذفه من النموذج . للمزيد من التفاصيل انظر Jeffrey M. Wooldridge, Introductory Econometrics: A Modem Approach, South-Western Publishing Co., 200, p. 386.

مفسر. قارن ذلك مع فرض اختبار Durbin-Watson والذي لا يسمح بوجود قيم للمتغير المنحدر عليه في فترات زمنية سابقة ضمن المتغيرات المنحدرة.

2 - كما سبق ولاحظنا، اختبار BG يمكن استخدامه حتى لو كان الخطأ يتبع نموذج متوسطات متحركة من الدرجة p، أي يكون حد الخطأ له النموذج التالي:

$$u_t = \varepsilon_t + \lambda_1 \varepsilon_{t-1} + \lambda_2 \varepsilon_{t-2} + \dots + \lambda_p \varepsilon_{t-p} \qquad (19.6.12)$$

حيث ε_t هو حد الخطأ الذي يستوفي كل الفروض التقليدية.

في فصول الاقتصاد القياسي المتعلق بالسلاسل الزمنية، سنستعرض نماذج الانحدار الذاتي والمتوسطات المتحركة من الدرجة p.

3 - إذا كان في (15.6.12) $p = 1$، فإن ذلك يعني ارتباطًا ذاتيًا من الدرجة الأولى، فإن اختبار BG يعرف باسم **اختبار Durbin's M**.

4 - من عيوب اختبار BG، أي القيمة p والتي تمثل طول الفترات الزمنية السابقة لا تستطيع تحديده مسبقًا. فلابد من عمل عدة تجارب لتحديدها. من الممكن أحيانًا الاعتماد على ما يسمى معلومات Akaike و Schwar Z لاختبار طول الفترة الزمنية السابقة.

Schwarz لاختبار طول الفترة الزمنية السابقة. سنناقش هذا الأسلوب في الفصل 13، ولاحقًا في هذا الفصل من خلال الاقتصاد القياسي للسلاسل الزمنية.

شرح اختبار BG: Illustration of BG Test

العلاقة بين الأجور - الإنتاجية The Wages-Productivity Relation

لشرح هذا الاختبار، سنطبقه على مثالنا التوضيحي باستخدام نموذج AR(6)، حصلنا على النتائج الموجودة في تمرين 25.12 من نتائج الانحدار المعطاة هناك، نجد أن $(n - p) = 34$ و $R^2 = 0.8920$. وبالتالي حاصل ضربهما يساوي قيمة كاي التربيعية المساوية لـ 30.328. وباستخدام درجات حرية = 6 (لماذا؟).

احتمال أن نصل على قيمة كاى التربيعية مساوية لـ 30.328 أو أكبر أو أصغر للغاية، جداول كاى التربيعية موجودة في ملحق D .4 يوضح أن احتمال الحصول على قيمة كاى التربيعية المساوية لـ 18.5476 أو أكبر يساوي 0.005. وبالتالي عند نفس المستوى من درجات الحرية، احتمال الحصول على قيمة كاى التربيعية المساوية لـ 30 ستكون صغيرة للغاية. في الواقع فإن القيمة الفعلية لـ p-value هي الصفر تقريبًا.

وبالتالي، فإننا نستنتج أنه في مثالنا الحالي، على الأقل واحد من معاملات الارتباط الذاتي الستة لا يساوى الصفر. باستخدام فترات زمنية سابقة من 1 إلى 6. نجد أنه فقط في النموذج AR(1) يكون المعامل معنويًا، مما يعني أنه لا داعي للاعتماد على أكثر من فترة زمنية واحدة فقط. مما يعني أن اختبار BG تحول إلى اختبار Durbin's m.

لماذا يوجد العديد من اختبارات الارتباط الذاتي؟
Why So Many Tests of Autocorrelation?

إجابة هذا السؤال كالتالي "... لا يوجد اختبار محدد يمكن وصفه بالاختبار الأفضل [أي له قوة أعلى بالمعنى الإحصائي] وبالتالي على المحلل أن يختار الاختبار المناسب للمشكلة محل الدراسة لاكتشاف وجود وشكل الارتباط الذاتي". [(34)] بالطبع نفس الاستنتاج والمقولة تم تناولهما في اختبارات اختلاف التباين التي استعرضناها في الفصل السابق.

7.12 ماذا تفعل عندما نجد ارتباطًا ذاتيًا: مقاييس إصلاحية:
WHAT TO DO WHEN YOU FIND AUTOCORRELATION: REMEDIAL MEASURES

إذا وجدنا ارتباطًا ذاتيًا في البيانات بعد تطبيق واحد أو أكثر من الاختبارات السابق ذكرها في الفقرة السابقة. ماذا نفعل بعد ذلك؟ لدينا أربعة خيارات:

1 - حاول أن تكتشف ما إذا كان هذا الارتباط الذاتي ارتباطًا ذاتيًا محضًا أم لا، فقد يكون موجودًا بسبب سوء توصيف النموذج.

فكما سبق وذكرنا في الفقرة 1.12، أحيانًا نلاحظ وجود نمط محدد في البواقي لأن النموذج - أو الشكل الدال المستخدم يكون غير سليم.

2 - إذا كان الارتباط الذاتي هو ارتباط محض وغير متعلق بأي مشكلة أخرى، من الممكن أن يستخدم الفرد تحويلة مناسبة للنموذج الأصلي، بحيث إن النموذج

---

(34) Ron C. Mittelhammer et al., op. cit., p. 547. تذكر أن قوة الاختبار الإحصائي هي 1 مطروح منه احتمال الخطأ من النوع الثاني، أي 1-احتمال قبول فرض خاطئ. أقصى قوة لأي اختبار هي 1 وأقل قيمة هي 0. كلما اقتربت قوة الاختبار من الصفر كلما زاد سوء هذا الاختبار وكلما اقترب من 1 كلما كان أكثر قوة. وما يقوله هؤلاء الكتاب هو أنه لا يوجد اختبار أكثر قوة من الآخرين في حالة الارتباط الذاتي.

المحول يكون خاليًا من مشكلة الارتباط الذاتي (الخالص). وكما هو الحال في مشكلة اختلاف التباين، لابد أن نستخدم نوعًا ما من طريقة المربعات الصغرى العامة (GLS).

3 - في حالة العينات كبيرة الحجم، من الممكن أن نستخدم طريقة Newey-West للحصول على الأخطاء المعيارية لمقدرات الـ OLS المرتبطة، وهذه الطريقة تعتبر امتدادًا لطريقة الأخطاء المعيارية المتسقة لـ White في حالة اختلاف التباين، وقد سبق واستعرضنا هذه الطريقة في الفصل السابق.

4 - في بعض الحالات يمكن أن نظل نستخدم طريقة OLS بصورة طبيعية.

ونظرًا لأهمية هذه الخيارات الأربعة، فقد خصصنا فقرة لكل خيار كالتالي:

8.12 خطأ توصيف النموذج في مقابل الارتباط الذاتي المحض: MODEL MIS-SPECIFICATION VERSUS PURE AUTOCORRELATION

بالعودة إلى انحدار الأجور - الإنتاجية المعطى في (1.5.12). رأينا أن قيمة d كانت 0.1229 ووفقًا لاختبار Durbin-Watson فإننا نستنتج أن هناك ارتباطًا طرديًا في حد الخطأ. هل قد يكون السبب في ظهور مثل هذا الارتباط كون النموذج الحالي غير سليم من ناحية التوصيف؟ بما أن البيانات الخاصة بالانحدار (1.5.12) هي بيانات سلاسل زمنية، من الممكن أن يكون هناك اتجاه عام في كل من الأجور أو الإنتاجية. إذا حدث ذلك، فإننا نحتاج إلى إدخال الزمن أو الاتجاه العام، t، في النموذج حتى نستطيع أن نرى العلاقة الخالصة بين الأجور والإنتاجية، بغض النظر عن الاتجاه العام الموجود في كل من المتغيرين.

لنرى ذلك، دعنا نتضمن متغير الاتجاه العام في (1.5.12) ونحصل على النتائج التالية:

$$\begin{aligned} \hat{Y}_t &= 1.4752 + 1.3057X_t - 0.9032t \\ se &= (13.18) \quad (0.2765) \quad (0.4203) \\ t &= (0.1119) \quad (4.7230) \quad (-2.1490) \\ & R^2 = 0.9631; \quad d = 0.2046 \end{aligned} \tag{1.8.12}$$

تفسير هذا النموذج يتم بطريقة مباشرة تمامًا: بمرور الزمن، مؤشر الأجر الحقيقي ينخفض بحوالي 0.9 وحدة لكل سنة. ووفقًا لذلك، إذا زاد مؤشر الإنتاجية بوحدة

واحدة، في المتوسط، فإن مؤشر الأجر الحقيقي سيرتفع بحوالي 1.3 وحدة، وعلى الرغم من أن هذا الرقم لا يختلف إحصائيًا عن الواحد (لماذا؟) إلا أنه يمكن ملاحظة إنه رغم إدخال متغير الزمن أو الاتجاه العام في النموذج، إلا أن قيمة d مازالت صغيرة جدًا، مما يعني أن (1.8.12) يعاني من ارتباط ذاتي محض، ولا يوجد خطأ توصيف النموذج.

كيف نعلم أن (1.8.12) هو التوصيف الصحيح للنموذج؟ لاختبار ذلك، دعنا نقوم بعمل انحدار لـ Y على X و X^2 لاختبار إمكانية أن يكون مؤشر الأجر الحقيقي غير خطي في علاقته بمؤشر الإنتاجية. النتائج الخاصة بهذا الانحدار جاءت كالتالي:

$$\begin{aligned} \hat{Y}_t &= -16.2181 + 1.9488X_t - 0.0079X_t^2 \\ t &= (-5.4891) \quad (24.9868) \quad (-15.9363) \\ & R^2 = 0.9947 \quad d = 1.02 \end{aligned} \tag{2.8.12}$$

هذه النتائج مهمة للغاية. فنحن نجد الآن أن كل المعاملات لها معنوية إحصائية عالية، قيم p-value صغيرة للغاية. من المقدار التربيعي السالب، نجد أنه على الرغم من أن مؤشر الأجر الحقيقي يزداد كلما ازداد مؤشر الإنتاجية، إلا أنه يزداد بمعدل متناقص. ولكن انظر الآن إلى قيمة d، مازالت تدل على وجود ارتباط ذاتي طردي في البواقي، حيث $d_t = 1.391$ و $d_u = 1.60$ وقيمة d المقدرة أقل من d_L.

من الممكن أن نستنتج من التحليل السابق أن انحدار الأجور والإنتاجية يعاني من ارتباط ذاتي محض، ولا يوجد أثر لخطأ توصيف النموذج، ولأننا على علم بعواقب الارتباط الذاتي، فإنه لابد من اتخاذ بعض الإجراءات التصحيحية، وسنقوم بعمل ذلك قريبًا.

في كل الانحدارات السابقة الخاصة بالأجور والإنتاجية التي استعرضناها من قبل، تم تطبيق اختبار الاعتيادية لـ Jarque-Bera ووجدنا أن البواقي تتبع التوزيع الطبيعي، حيث إننا نحتاج إلى هذا الفرض لتطبيق اختبار d على حد الخطأ.

9.12 تصحيح الارتباط الذاتي (المحض): طريقة المربعات الصغرى العامة (GLS): CORRECTING FOR (PURE) AUTOCORRELATION: THE METHOD OF GENERALIZED LEAST SQUARES (GLS)

نظرًا لعواقب الارتباط الذاتي، وخصوصًا عدم كفاءة مقدرات الـ OLS، نحتاج إلى تصحيح أو علاج هذه المشكلة. هذا العلاج يعتمد على المعلومات المتاحة عن طبيعة العلاقة التبعية في حد الخطأ أو بالأحرى معرفة شكل الارتباط الذاتي.

كبداية، دعنا نعتبر النموذج ثنائي المتغيرات التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1.9.12)$$

وافترض أن حد الخطأ يتبع AR(1) كالتالي:

$$u_t = \rho u_{t-1} + \varepsilon_t \quad -1 < \rho < 1 \qquad (2.9.12)$$

والآن لدينا حالتان: (1) ρ معلومة و(2) ρ غير معلومة ولكن يمكن تقديرها.

عندما تكون ρ معلومة: When ρ is Known

إذا كان معامل الارتباط الذاتي من الدرجة الأولى معلومًا، يمكن حل مشكلة الارتباط الذاتي بسهولة. فإذا تحققت (1.9.12) عند الزمن t، فإنها تتحقق أيضًا عند $(t - 1)$ وبالتالي نجد أن

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + u_{t-1} \qquad (3.9.12)$$

وبضرب (3.9.12) في ρ من الطرفين نحصل على التالي:

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1} \qquad (4.9.12)$$

وبطرح (4.9.12) من (1.9.12) نحصل على:

$$(Y_t - \rho Y_{t-1}) = \beta_1(1 - \rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t \qquad (5.9.12)$$

حيث $\varepsilon_t = (u_t - \rho u_{t-1})$

ويمكن التعبير عن (5.9.12) كالتالي:

$$Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t \qquad (6.9.12)$$

حيث $\beta_1^* = \beta_1\ (1-\rho)$ و $Y_t^* = (Y_t - \rho Y_{t-1})$ و $X_t^* = (X_t - \rho X_{t-1})$ و $\beta_2^* = \beta_2$ بما أن مقدار الخطأ في (6.9.12) مستوفي كل فروض OLS على المتغيرات المحولة Y^* و X^* ونحصل على مقدرات لها كل الخصائص الجيدة المعروفة، أي BLUE.

وهنا يكون تطبيق (6.9.12) مناظرًا لاستخدام المربعات الصغرى العامة (GLS) التي سبق واستعرضناها في الفصل السابق - تذكر أن GLS ما هي إلا OLS مطبقة على النموذج المحول والذي أصبح بعد التحويل مستوفيًا كل الفروض التقليدية.

انحدار (5.9.12) معروف باسم المعادلة العامة أو معادلة الفروق. هو يعتمد على انحدار y على x، ليس في صورتهما الطبيعية، ولكن في صورة الفروق والتي يتم الحصول عليها بعد طرح نسبة (p=) من قيمة المتغير من قيمته في فترة زمنية سابقة للفترة الحالية. في عملية الفروق نفقد مشاهدة واحدة، حيث إن المفردة الأولى لا يوجد ما يسبقها. لعلاج هذا الفقد، المشاهدة الأولى من Y و X يتم تحويلها كالتالي:[35] $Y_1\sqrt{1-\rho^2}$ و $X_1\sqrt{1-\rho^2}$ وهذه التحويلة معروفة باسم تحويلة Prais-Winsten.

عندما تكون ρ غير معلومة : When ρ is not Known

على الرغم من أن الطريقة السابق ذكرها مباشرة تمامًا، إلا أن تطبيقها في الواقع صعب، حيث يصعب عمليًا أن تكون ρ معلومة. وبالتالي لابد أن نجد طريقة لتقدير ρ. وفي هذا الاطار لدينا العديد من الخيارات كالتالي:

طريقة الفروق الأولى : The First-Difference Method

بما أن ρ تقع بين 0 و ±1، يمكن أن نبدأ بهاتين القيمتين، فمن الممكن افتراض الحد الأدنى، وتكون $\rho = 0$، أي لا يوجد ارتباط تسلسلي (من الدرجة الأولى) والافتراض الآخر يكون باستخدام الحد الأعلى، أي تكون $\rho = \pm 1$، أي أن هناك ارتباطًا تامًا طرديًا أو عكسيًا.

في واقع الأمر، إذا قمنا بعمل انحدار، فإننا نفترض أنه لا يوجد ارتباط ذاتي، ثم باستخدام اختبار Durbin-Watson أو أي اختبار آخر نحكم على مدى صحة هذا

(35) فقدان مشاهدة واحدة قد لا يكون له أهمية كبيرة خصوصًا في العينات كبيرة الحجم، ولكن قد يكون له تأثير واضح في حالة العينات صغيرة الحجم. بدون تحويل المشاهدة الأولى كما ذكرنا، سيكون تباين الخطأ غير ثابت. لمزيد من التفاصيل انظر

Jeffrey Wooldridge, op. cit., p. 388.

لبعض نتائج محاكاة Monte Carlo والخاصة بأهمية المشاهدة الأولى انظر Russell Davidson and James G. Mackinnon Estimation and Inference in Econometrics, Oxford University Press, New York, 1993, Table 10.1, p. 349.

الفرض. فاذا كانت $\rho = \pm 1$، فإن معادلة الفروق العامة (5.9.12) تتقلص وتأخذ شكل معادلة الفروق الأولى التالية:

$$Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$$

أو

$$\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t \qquad \textbf{(7.9.12)}$$

حيث Δ تمثل مؤشر الفروق الأولى والذي استعرضناه في (10.1.12) بما أن الخطأ في (7.9.12) لا يوجد فيه ارتباط تسلسلي من الدرجة الأولى (لماذا؟)، فإننا يمكن أن نقوم بعمل انحدار (7.9.12) عن طريق الحصول على الفروق الأولى لكل من المتغيرين المنحدر والمنحدر عليه، ثم تطبيق الانحدار على هذه الفروق الأولى.

تحويلة الفروق الأولى يمكن استخدامها عندما يكون معامل الارتباط الذاتي كبيراً جداً، أي يزيد عن 0.8، أو قيمة d لـ Durbin-Watson صغيرة للغاية. Maddala قدم لنا القاعدة التالية: استخدم الفروق الأولى عندما تكون $d < R^2$.[36] وهذه هي الحالة في انحدار الأجور والإنتاجية الموجود في (1.5.12) والذي كانت قيمة $d = 0.1229$ و $r^2 = 0.9584$.

انحدار الفروق الأولى لمثالنا التوضيحي سنستعرضها بعد قليل.

والخاصية المميزة لنموذج الفروق الأولى (7.9.12) أنها خالية من الجزء الثابت المقطوع من المحور الصادي. وبالتالي لتقدير (7.9.12) نستخدم الانحدار المار بنقطة الأصل (أي بدون جزء ثابت مقطوع من المحور الصادي) وتقدير هذا الانحدار متاح الآن في غالبية حزم الحاسب الآلي الإحصائية. إذاً، لأي سبب من الأسباب، نسيت أن تسقط الجزء الثابت المقطوع من المحور الصادي، وقدرت النموذج الذي يشتمل عليه كالتالي:

$$\Delta Y_t = \beta_1 + \beta_2 \Delta X_t + \varepsilon_t \qquad \textbf{(8.9.12)}$$

فإن النموذج الأصلي لابد وأن يشتمل على اتجاه عام، و β_1 ستمثل معامل متغير الاتجاه العام.[37] وبالتالي من المميزات «المفاجئة» لوجود جزء ثابت مقطوع من المحور

(36) Maddala, op. cit., p. 232.

(37) من السهل توضيح ذلك، دع $Y_t = \alpha_1 + \beta_{1t} + \beta_2 X_t + u_t$. وبالتالي $Y_{t-1} = \alpha + \beta_1(t-1) + \beta_2 X_{t-1} + u_{t-1}$ والذي يوجد فيه جزء ثابت مقطوع من المحور الصادي وهو عبارة عن معامل متغير الاتجاه العام الموجود في النموذج الأصلي. تذكر أننا نفترض أن $\rho = 1$.

الصادي في نموذج الفروق الأولى هو إمكانية اختبار معنوية وجود متغير الاتجاه العام في النموذج الأصلي.

بالعودة إلى انحدار الأجور – الإنتاجية (1.5.12)، وباستخدام AR(1) وقيمة لـ d صغيرة بالنسبة لـ r^2، نعيد عمل انحدار (1.5.12) باستخدام شكل الفروق الأولى بدون جزء ثابت مقطوع من المحور الصادي، تذكر أن (1.5.12) هو المستوى الأساسي فستكون النتائج كالتالي:[38]

$$\widehat{\Delta Y_t} = 0.7199 \Delta X_t$$
$$t = (9.2073) \qquad r^2 = 0.3610 \qquad d = 1.5096 \qquad \textbf{(9.9.12)}$$

بمقارنة ذلك مع انحدار المستوى الأساسي (1.5.12)، نرى أن معامل الميل لم يتغير كثيراً، ولكن قيمة r^2 انخفضت بشكل ملحوظ، وهذا عموماً حدث نتيجة استخدام الفروق الأولى، حيث إننا ندرس سلوك المتغيرات حول قيم الاتجاه العام (الخطية). بالطبع لا نستطيع مقارنة r^2 الخاصة بـ (9.9.12) مباشرةً مع r^2 الخاصة بـ (1.5.12) لأن المتغيرات التابعة في النموذجين مختلفة.[39] لاحظ أيضاً أنه بمقارنة مع النموذج الأصلي، نجد أن قيمة d زادت بشكل ملحوظ، مما قد يعني أن هناك مقداراً ما من الارتباط الذاتي في انحدار الفروق الأولى.[40]

ملحوظة أخرى مثيرة للاهتمام في تحويلة الفروق الأولى مرتبطة بخاصية السكون للسلسلة الزمنية محل الدراسة. بالعودة إلى المعادلة (1.2.12) والتي تصف نموذج AR(1)، لاحظ أنه إذا كان في الواقع $\rho = 1$، فإنه يتضح من المعادلات (3.2.12) و (4.2.12) أن لسلسلة u_t غير ساكنة حيث التباين والتغاير أصبحا غير محدودين. وهذا هو السبب الذي جعلنا نفترض أن $|\rho| < 1$ عند مناقشة هذا الموضوع. ولكن يتضح من (1.2.12) تصبح كالتالي:

---

(38) في تمرين 38.12 مطلوب منك عمل هذا الانحدار مع إضافة الجزء الثابت.

(39) هناك أيضاً ملاحظة تتعلق بمقارنة r^2 الخاصة بالنموذج الأساسي r^2 الخاصة بالفروق الأولى، للمزيد من التفاصيل الخاصة بذلك انظر .Maddala, op. cit., Chap. 6

(40) ليس من الواضح أو المؤكد ما إذا كانت d المحسوبة من انحدار الفروق الأولى ممكن تفسيرها بنفس طريقة النموذج الأصلي للانحدار. عموماً باستخدام اختبار الدفعات ممكن ملاحظة أنه لا يوجد دليل على الارتباط الذاتي في البواقي في نموذج الفروق الأولى.

$$u_t = u_{t-1} + \varepsilon_t$$

أو

$$(u_t - u_{t-1}) = \Delta u_t = \varepsilon_t \qquad (10.9.12)$$

أي أن u_t للفروق الأولى ساكن، ويساوي ε_t والذي يعتبر حد خطأ عشوائي ينطبق عليه الفروض التقليدية.

الهدف من المناقشة السابقة، أنه إذا كانت السلسلة الزمنية الأصلية غير ساكنة، فإنه عادة ما يؤدي الحصول على فروقها الأولى إلى تسكينها. وبالتالي تحويلة الفروق الأولى لها منفعة مزدوجة، حيث إنها قد تؤدي إلى التخلص من الارتباط الذاتي (من الدرجة الأولى) بالإضافة إلى جعل السلسلة الزمنية ساكنة. سنعود مرة أخرى إلى هذا الموضوع في **الجزء V**، والذي سنناقش فيه بمزيد من التفاصيل تحليل السلاسل الزمنية في الاقتصاد القياسي.

ذكرنا أن تحويلة الفروق الأولى قد يفضل استخدامها عندما تكون ρ كبيرة أو d صغيرة. وبشكل أكثر تحديداً، تحويلة الفروق الأولى تكون سليمة فقط إذا كانت $\rho = 1$. وفي واقع الأمر، هناك اختبار يسمى **اختبار Berenblutt-Webb test**[41]، لاختبار الفرض الخاص بـ $\rho = 1$. وإحصاء الاختبار يسمى **إحصاء g**، والذي يعرف كالتالي:

$$g = \frac{\sum_2^n \hat{e}_t^2}{\sum_1^n \hat{u}_t^2} \qquad (11.9.12)$$

حيث $\hat{u}_t$ هو بواقي OLS من الانحدار الأصلي (المستوى الأساسي) و e_t هو بواقي الـ OLS من نموذج الفروق الأولى. مع الوضع في الاعتبار أن شكل الفروق الأولى لا يشتمل على جزء ثابت مقطوع من المحور الصادي.

لاختبار معنوية إحصاء g، وبافتراض أن مستوى الانحدار الأساسي يشتمل على جزء ثابت مقطوع من المحور الصادي، يمكن استخدام جداول Durbin-Watson والفرق الوحيد هو أن الفرض العدمي الحالي هو $\rho = 1$ وليس فرض Durbin-Watson التقليدي وهو $\rho = 0$.

(41) I. I. Berenblutt and G. I. Webb, "A New Test for Autocorrelated Errors in the Linear Regression Model," Journal of the Royal Statistical Society, Series B, vol. 35, No. 1, 1973, pp. 33–50.

بالعودة إلى انحدار الأجور - الإنتاجية، باستخدام نموذج الانحدار الأصلي (1.5.12) حصلنا على $\sum\hat{u}_t^2 = 272.0220$ أما الانحدار الأول (11.7.12) فحصلنا على $\sum\hat{e}_t^2 = 0.334270$. وبالتعريض عن هاتين القيمتين في إحصاء g المعطى في (11.9.12) نحصل على :

$$g = \frac{33.4270}{272.0220} = 0.0012 \qquad \textbf{(12.9.12)}$$

وبناء على جدول Durbin-Watson وباستخدام 39 مفردة ومتغيراً مفسراً واحداً. نجد أن $d_L = 1.435$ و $d_u = 1.540$ (5% مستوى معنوية). بما أن g أقل من القيمة الصغرى (d) لا نرفض الفرض العدمي القائل بأن p الحقيقة = 1. ضع في الاعتبار أنه على الرغم من أننا نستخدم نفس جدول Durbin-Watson إلا أن الفرض العدمي الحالي هو $\rho = 1$ وليس $\rho = 0$. وفقًا لهذه القيم، فإن النتائج المعطاة في (9.9.12) تكون مقبولة.

ρ بناء على إحصاء d لـ Durbin-Watson؛ ρ Based on Durbin-Watson d Statistic

إذا لم نستطع استخدام تحويلة الفروق الأولى، لأن ρ ليست قريبة معنويًا من الوحدة، فإن هناك طريقة سهلة لتقديرها من العلاقة بين d و ρ السابق ذكرها في (10.6.12). وبالتالي يمكننا تقدير ρ كالتالي :

$$\hat{\rho} \approx 1 - \frac{d}{2} \qquad \textbf{(13.9.12)}$$

أي أنه في أحجام العينات الكبيرة نسبيًا يمكن الحصول على تقدير rho من (13.9.12) واستخدامه لتحويل البيانات كما هو موضح في معادلة الفروق العامة (5.9.12). ولكن ضع في الاعتبار أن العلاقة المذكورة بين ρ و d والمعطاة في (13.9.12) قد لا تكون سليمة في العينات صغيرة الحجم. وقد قدم Theil و Nagar تعديلاً مناسبًا لمثل هذه الحالة. هذا التعديل معطى في تمرين 6.12.

في انحدارنا الخاص بالأجور - الإنتاجية (1.5.12)، حصلنا على قيمة d مساوية لـ 0.1229. باستخدام هذه القيمة في (13.9.12)، نحصل على $\hat{\rho} \approx 0.9386$. وباستخدام هذه القيمة المقدرة لـ rho، يمكننا تقدير انحدار (5.9.12). كل ما علينا فعله هو طرح 0.9386 مضروبة في قيمة Y في الفترة الزمنية السابقة من قيمتها الحالية، وبالمثل طرح 0.9386 مضروبة في قيمة X في الفترة الزمنية السابقة من قيمتها الحالية ثم عمل

انحدار OLS على المتغيرات المحولة كما في (6.9.12) حيث $Y_t^* = (Y_t - 0.9386Y_{t-1})$ و $X_t^* = (X_t - 0.9386X_{t-1})$.

ρ المقدرة من البواقي: **ρ Estimated from the Residuals**

إذا كان من الممكن استخدام AR(1) لـ $u_t = \rho u_{t-1} + \varepsilon_t$، هناك طريقة بسيطة لتقدير rho وهي تعتمد على عمل انحدار للبواقي $\hat{u}_t$ على $\hat{u}_{t-1}$. حيث إن $\hat{u}_t$ هي مقدرات متسقة لقيم u_t الحقيقية، كما سبق وذكرنا، أي أننا نقوم بعمل الانحدار التالي:

$$\hat{u}_t = \rho \,.\hat{u}_{t-1} + v_t \qquad (14.9.12)$$

حيث $\hat{u}_t$ هي البواقي التي نحصل عليها من الانحدار الأصلي (المستوى الأساسي) و v_t هو حد الخطأ في هذا النموذج. لاحظ أنه لا يوجد داع لأن يشمل النموذج (14.9.12) على جزء ثابت مقطوع من المحور الصادي، حيث إننا نعلم أن بواقي OLS تجمع إلى صفر.

بواقي نموذجنا الخاص بالأجور - الإنتاجية المعطى في (1.5.12) سبق عرضه في جدول (5.12). باستخدام هذه البواقي، حصلنا على نتائج الانحدار التالية:

$$\begin{aligned} \widehat{\hat{u}_t} &= 0.9142\hat{u}_{t-1} \\ t &= (16.2281) \qquad r^2 = 0.8736 \end{aligned} \qquad (15.9.12)$$

وكما يوضح الانحدار، $\hat{\rho} = 0.9142$. باستخدام هذا التقدير، يمكن تحويل النموذج الأصلي كما في (6.9.12). وحيث إن القيمة المقدرة لـ rho مساوية تقريبًا للقيمة التي حصلنا عليها من Durbin-Watson d فإن نتائج الانحدار باستخدام rho (15.9.12) يجب ألا تختلف كثيرًا عن نظيرها الذي تم الحصول عليه من تقدير rho بناء على Durbin-Watson d. وسنترك للقارئ إثبات ذلك.

الطرق التكرارية لتقدير ρ: **Iterative Methods of Estimating ρ**

كل طرق تقدير ρ السابق ذكرها، تعطي قيمة تقديرية واحدة لـ ρ. ولكن هناك ما يُسمى بالطرق التكرارية لتقدير ρ بشكل متكرر، أي بالتقريب، وتبدأ بقيمة مبدئية ρ. من أمثلة هذه الطرق: **الطريقة التكرارية لـ Cochrane-Orcutt و طريقة الخطوتين لـ Cochrane-Orcutt وطريقة الخطوتين لـ Burbin وطريقة البحث لـ Hildreth-Lu** .

بالطبع أكثر هذه الطرق شهرة الطريقة التكرارية لـ Cochrane-orcutt الطرق التكرارية تم استعراضها في إطار هذا الكتاب من خلال عدد من التمارين. تذكر أن الهدف الأسمى لهذه الطرق هو تقديم تقدير لـ ρ يمكن استخدامه للحصول على مقدرات GLS للمعالم المجهولة. إحدى مميزات استخدام الطريقة التكرارية لـ Cochrane-Orcutt أنها يمكن استخدامها ليس فقط لتقدير عملية AR(1) ولكن تستخدم أيضًا مع عمليات انحدار ذاتي من رتب أعلى، مثل $\hat{u}_t = \hat{\rho}_1\hat{u}_{t-1} + \hat{\rho}_2\hat{u}_{t-2} + v_t$ والتي تعتبر AR(2). فبالحصول على القيمتين الخاصتين بـ rho، يمكن استخدام معادلة الفروق العامة (6.9.12) إذا كان هناك أيضًا أكثر من أثنين rho. والحاسب الآلي يمكنه الآن القيام بذلك بسهولة.

بالعودة إلى نموذج الأجور – الإنتاجية، وبافتراض العملية AR(1)، وباستخدام الطريقة التكرارية لـ Cochrane-Orcutt، نحصل على القيم التالية للـ rho: 0.9142، 0.9052، 0.8992، 0.8956، 0.8935، 0.8924 و 0.8919. القيمة الأخيرة 0.8919 يمكن استخدامها لتحويل النموذج الأصلي كما في (6.9.12) وتقديره باستخدام الـ OLS. بالطبع OLS على النموذج المحول هي ببساطة GLS. النتائج كالتالي:

$$\begin{aligned} \hat{Y}_t^* &= 45.105 \;+\; 0.5503X_t^* \\ \text{se} &= (6.190) \quad (0.0652) \\ t &= (7.287) \quad (8.433) \qquad r^2 = 0.9959 \end{aligned} \tag{16.9.12}$$

بمقارنة نتائج الانحدار مع النموذج الأصلي المعطى في (1.5.12) نرى أن معامل الميل قد انخفض بشكل ملحوظ، ولاحظ أيضًا شيئين: الأول، معامل الجزء الثابت في (16.9.12) هو $\beta_1(1-\rho)$ والذي يمكن استخدامه للحصول على β_1 بسهولة حيث إننا نعلم أن $\rho = 08913$.

ثانيًا، r^2's للنموذج المحول (16.9.12) وللنموذج الأصلي (1.5.12) لا يمكن مقارنتهما مباشرةً، حيث إن المتغيرات التابعة في النموذجين مختلفة.

الاحتفاظ بالمشاهدة الأولى : Retaining the First Observation

سبق وذكرنا أنه في العينات صغيرة الحجم يكون مؤثر الاحتفاظ بالمشاهدة الأولى أو حذفها فقط يؤدي ذلك إلى فرق كبير، وطبعًا في العينات كبيرة الحجم قد يكون هذا الفرق غير مؤثر.

الاحتفاظ بالمشاهدة الأولى لـ à la Prais-Winsten، نحصل على نتائج الانحدار التالية :[42]

$$\hat{Y}_t^* = 26.454 \quad + \quad 0.7245X_t^*$$
$$\text{se} = (5.4520) \qquad (0.0612) \qquad\qquad (17.9.12)$$
$$t = (4.8521) \qquad (11.8382) \qquad r^2 = 0.9949$$

الفرق بين (16.9.12) و (17.9.12) يعطينا انطباعًا عن أهمية الاحتفاظ أو إسقاط المشاهدة الأولى وما يمكن ان نسميه كفرق جوهري بين نتائج الانحدارين. لاحظ أيضًا أن معامل الميل في (17.9.12) تقريبًا متساويًا مع نظيره في (1.5.12).

تعليقات عامة: General Comments

هناك العديد من النقاط الواجب الإشارة إليها والخاصة بتصحيح الارتباط الذاتي وفقًا للطرق المختلفة العديدة السابق ذكرها:

أولاً، بما أن مقدرات OLS تعتبر مقدرات متسقة بغض النظر عن الارتباط الذاتي في حالة العينات كبيرة الحجم، فإن هناك فرقًا بسيطًا بين تقدير قيمة ρ من Durbin-Watson d أو من انحدار بواقي الفترة الزمنية الحالية على بواقي الفترة السابقة أو من العملية التكرارية لـ Cochrane-Orcutt وذلك لأنها جميعًا تقدم مقدرات متسقة لقيمة ρ الحقيقية.

ثانيًا، كل الطرق السابق ذكرها أعلى هي في الأساس طرق تتم على خطوتين، في الخطوة 1 نحصل على مقدر للمعلمة المجهولة ρ ، وفي الخطوة 2 يستخدم هذا التقدير لتحويل المتغيرات لتقدير معادلة الفروق العامة، والتي تعتبر في الأساس GLS. وبما أننا نستخدم $\hat{\rho}$ بدلاً من قيمة ρ الحقيقية، فإن كل هذه طرق من طرق التقدير تعرف باسم GLS الممكنة (FGLS) أو طرق GLS المقدرة (EGLS).

ثالثًا، من المهم ملاحظة أنه عندما نستخدم طريقة FGLS أو EGLS لتقدير معالم النموذج المحول، المعاملات المقدرة ليست بالضرورة تتمتع بكل الخصائص المميزة لها في حالة النموذج التقليدي، مثل BLUE، خصوصًا في العينات صغيرة الحجم.

(42) بإدخال المشاهدة الأولى ، القيم التكرارية لـ rho هي : 0.9142, 9.9462, 0.9556, 0.9591, 0.9605 and 0.9610 . القيمة الأخيرة استخدمت في البيانات المحولة من معادلة الفروق العامة .

وبدون الطرق إلى تقنيات معقدة، يمكن القول أنه كمبدأ عام، عندما نستخدم مقدر مكان القيمة الحقيقية، فإن مقدرات OLS للمعاملات قد تكون لها الخصائص المميزة بشكل تقاربي، أي في العينات كبيرة الحجم. وأيضًا كل العمليات المرتبطة باختبارات الفروض بوجه عام تكون تقاربية. في العينات صغيرة الحجم، يجب أن يكون الباحث شديد الحذر في تفسيره للنتائج التي يحصل عليها.

رابعًا، في حالة استخدام EGLS، إذا لم نحتفظ بالمشاهدة الأولى (كما هو الحال في عملية Cochrane-Orcutt)، فإنه قد تتأثر القيم الرقمية للمقدرات، وليس ذلك فقط دائمًا مدى كفاءة هذه المقدرات قد تتأثر كثيرًا خصوصًا إذا كانت أحجام العينات صغيرة، وإذا كانت المتغيرات المنحدرة غير عشوائية.[43] وبالتالي في العينات صغيرة الحجم من المهم الاحتفاظ بالمشاهدة الأولى à la Prais-Winsten . بالطبع، إذا كان الحجم كبيرًا نسبيًا وسواء بوجود أو عدم وجود المشاهدة الأولى، فإن EGLS تعطي نتائج متشابهة. ومن الطريف ملاحظة أن EGLS مع تحويلة Prais-Winsten معروفة باسم EGLS الكاملة أو للاختصار FEGLS.

10.12 طريقة Newey-West لتصحيح الأخطاء القياسية للـ OLS: THE NEWEY–WEST METHOD OF CORRECTING THE OLS STANDARD ERRORS

بدلاً من استخدام طرق FGLS السابق مناقشتها في الفقرة السابقة، مازال من الممكن استخدام OLS ، ولكن مع تصحيح الأخطاء القياسية للارتباط الذاتي وفقًا للطريقة التي اقترحها West و Newey.[44] وهذه الطريقة تعتبر امتدادًا لطريقة الأخطاء القياسية المتسقة لـ White في حالة اختلاف التباين والتي سبق وناقشناها في الفصل السابق.

الأخطاء القياسية المصححة معروفة باسم الأخطاء القياسية HAC (اختلاف التباين والارتباط الذاتي - المتسق).

---

(43) وهذا الأمر يغلب حدوثه إذا كانت المتغيرات المنحدرة بها اتجاه عام . وهذا هو الغالب في الكثير من البيانات الاقتصادية .

(44) W. K. Newey, and K. West, "A Simple Positive Semi-Definite Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, Econometrica, vol. 55, 1987, pp. 703–708.

أو باختصاف معروفة باسم الأخطاء القياسية لـ Newey-West.

لن نستعرض الخلفية الرياضية لطريقة Newey-West لما تشمله من بعض التفاصيل والتعقيدات الرياضية.[45]

لكن غالبية الحزم الإلكترونية الحديثة يمكنك استخدامها لحساب الأخطاء القياسية لـ Newey-West. ولكن الشيء المهم والذي نود الإشارة إليه هنا هو أن عملية Newey-West لا يمكن تطبيقها إلا في العينات كبيرة الحجم، وقد لا يمكن تطبيقها في العينات صغيرة الحجم. ولكن في الأحجام الكبيرة لدينا الآن طريقة تولد أخطاء قياسية مصححة من الارتباط الذاتي، وبالتالي لا داعي بأن نهتم بتحويلة EGLS السابق ذكرها في الفصل السابق. وبالتالي إذا كان حجم العينة كبيراً بدرجة كافية، من الأفضل للباحث أن يستخدم طريقة Newey-West لتصحيح الأخطاء القياسية للـ OLS ليس فقط في حالة وجود ارتباط ذاتي، ولكن أيضًا في حالة اختلاف التباين، حيث إن طريقة HAC تعلاج كلاً من المشكلتين، أما طريقة White فإنها خاصة فقط بمشكلة اختلاف التباين.

مرة أخرى، دعنا نعيد استخدام انحدار الأجور - الإنتاجية الموجود في (1.5.12). نحن نعلم أن هذا الانحدار يعاني من الارتباط الذاتي.

حجم العينة في هذا المثال يساوي 40 مشاهدة، وهذا الحجم يعتبر كبيراً نوعًا ما، وبالتالي يمكن أن تستخدم طريقة HAC. باستخدام Eviews 4 نحصل على نتائج الانحدار التالية:

$$
\begin{aligned}
\hat{Y}_t &= 29.5192 \quad + \quad 0.7136\hat{X}_t \\
\text{se} &= (4.1180)^* \quad (0.0512)^* \qquad\qquad (1.10.12) \\
& \qquad\qquad r^2 = 0.9584 \qquad d = 0.1229
\end{aligned}
$$

حيث * تعني الأخطاء القياسية HAC.

بمقارنة هذا الانحدار مع (1.5.12)، نجد أن كلاً من معادلات المعاملات المقدرة وقيمة r^2 متساوية. ولكن لاحظ أن الأخطاء القياسية HAC أكبر من الأخطاء القياسية للـ OLS وبالتالي نسب t وفقًا للـ HAC أصغر كثيراً من نسب t وفقًا للـ OLS. وهذا

(45) إذا كنت تستطيع التعامل مع جبر المصفوفات، هذه الطريقة مشروحة في Greene, op. cit, 4th ed., pp. 462–463.

يوضح أن OLS قد قدرت الأخطاء القياسية الحقيقية بأقل من قيمتها الفعلية . والمثير للانتباه هنا أن إحصاء d في كل من (1.5.12) و (1.10.12) متساويان . ولكن لا تقلق حيث إن طريقة HAC قد أخذت في اعتبارها تصحيح الأخطاء القياسية للـ OLS .

11.12 OLS مقابل FGLS و HAC :

المشكلة العملية التي تواجه الباحث الآن : أنه في حالة وجود ارتباط ذاتي ، فإن مقدرات OLS على الرغم من أنها غير متحيزة ، متسقة وتقاربيًا تتبع التوزيع الطبيعي إلا أنها ليست كفئًا . وبالتالي عملية الاستدلال التقليدية بناء على اختبارات t، F وكاى التربيعي لم يعد من الممكن استخدامها . على الجانب الآخر ، مقدرات FGS و HAC تعتبر مقدرات كفئًا ، ولكن في حالة العينات المحدودة أو صغيرة الحجم ، فإن خصائص هذه المقدرات غير مضمونة على الإطلاق . فهذا يعني أنه في العينات صغيرة الحجم فإن FGLS و HAC قد تكون أسوأ من OLS . في واقع الأمر وبناء على دراسة محاكاة Monte Carlo قام بها كل من Griliches و Rao [46] فقد وجدا أنه إذا كان حجم العينة صغيرًا نسبيًا ومعامل الارتباط الذاتي ، ρ، أقل من 0.3 فإن OLS تعتبر أفضل أو مساوية لجودة FGLS .

وبشكل عملي ، فإنه من الممكن استخدام OLS في العينات صغيرة الحجم ، والتي يكون تقدير rho فيها أقل من 0.3 . بالطبع السؤال الخاص بمتى يمكن القول بأن العينة صغيرة الحجم أو كبيرة الحجم يعتبر سؤالاً نسبيًا ، ومن الأفضل أن يستخدم الباحث بعض الأساليب العملية للحكم . فإذا كان لديك فقط 15 إلى 20 مشاهدة ، فإن العينة تعتبر صغيرة ، أما إذا كان لديك ، مثلاً ، 50 أو أكثر من المشاهدات فإن العينة يمكن اعتبارها كبيرة الحجم .

12.12 التنبؤ وفقًا لحدود الأخطاء المترابطة ذاتيًا :
FORECASTING WITH AUTOCORRELATED ERROR TERMS

في الفقرة 10.5 ، قدمنا مبادئ التنبؤ في إطار نماذج لانحدار ثنائية المتغيرات باستخدام الإطار التقليدي . كيف يمكن استخدام هذه الأساسيات إذا وجد ارتباط

(46) Z. Griliches, and P. Rao, "Small Sample Properties of Several Two-stage Regression Methods in the Context of Autocorrelated Errors," Journal of the American Statistical Association, vol. 64, 1969, pp. 253–272.

ذاتي؟ وعلى الرغم من أن هذا الموضوع تمت مناقشته منفصلاً في المواد الدراسية المتعلقة بالتنبؤ الاقتصادي، إلا أنه يمكن أن نلقي عليه بعض الضوء في الكتاب الحالي.

ولمزيد من التحديد، سنظل مستخدمين النموذج ثنائي المتغيرات، ونفترض العملية AR(1) وبالتالي لدينا:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \qquad (1.12.12)$$

$$u_t = \rho u_{t-1} + \varepsilon_t \qquad -1 < \rho < 1 \qquad (2.12.12)$$

حيث ε_t هو مقدار الخطأ العشوائي.

بالتعويض عن (2.12.12) في (1.12.12) نحصل على:

$$Y_t = \beta_1 + \beta_2 X_t + \rho u_{t-1} + \varepsilon_t \qquad (3.12.12)$$

إذا أردت أن تتنبأ بـ Y للفترة الزمنية القادمة $(t+1)$، فسنحصل على:

$$Y_{t+1} = \beta_1 + \beta_2 X_{t+1} + \rho u_t + \varepsilon_{t+1} \qquad (4.12.12)$$

وبالتالي، فالقيمة المتنبأ بها للفترة الزمنية القادمة تتكون من ثلاثة أجزاء:

(1) القيمة المتوقعة = $(\beta_1 + \beta_2 X_{t+1})$، (2) ρ مضروبة في حد الخطأ في الفترة الزمنية السابقة (3) مقدار خطأ عشوائي تقليدي قيمته المتوقعة تساوي الصفر. بمعلومية قيمة X_{t+1}، يمكن تقدير (1) عن طريق $\hat{\beta}_1 + \hat{\beta}X_{t+1}$، حيث مقدرات OLS تم الحصول عليها من العينة ونقدر (2) وفقًا لـ $\hat{\rho}\hat{u}_t$ حيث $\hat{\rho}$ مقدرة بأحد الطرق السابق ذكرها في الفقرة 9.12. في الزمن $(t+1)$، قيمة $\hat{u}_t$ تعتبر معروفة بالفعل، وبالتالي القيمة المقدرة لـ Y_{t+1} في (4.12.12) هي:

$$\hat{Y}_{t+1} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+1} + \hat{\rho}\hat{u}_t \qquad (5.12.12)$$

وبنفس المنطق فإن

$$\hat{Y}_{t+2} = \hat{\beta}_1 + \hat{\beta}_2 X_{t+2} + \hat{\rho}^2\hat{u}_t \qquad (6.12.12)$$

للفترة الثانية وهكذا.

التنبؤ الذي قمنا به في الفقرة 5.12 يسمى التنبؤ الإستاتيكي أما الموجود في (5.12.12) و (6.12.12) فإنه يسمى التنبؤ الديناميكي، حيث إنه حتى تقوم بهذا التنبؤ

لابد من الأخذ في الاعتبار الأخطاء الموجودة في التنبؤ السابق. كما في الفقرة 10.15 سنحتاج إلى التنبؤ بالأخطاء (القياسية) لـ (5.12.12) و (6.12.12)، ولكن تصبح المعادلات أكثر تعقيداً. [47]

وبما أن **معظم الحزم الحديثة للاقتصاد القياسي**، مثلاً Microfit، Eviews أو Shazam يمكن من خلالها حساب التنبؤ بالأخطاء القياسية، فلا يوجد أي داع لاستعراض هذه المعادلات المعقدة في إطار الكتاب الحالي.

كمثال توضيحي، دعنا نعود إلى انحدار الأجور - الإنتاجية. تذكر أن بيانات العينة الخاصة بهذا المثال تقع في الفترة الزمنية 1959 إلى 1998.

وقد قمنا بإعادة تقدير النموذج باستخدام بيانات 1959-1996 فقط واحتفظنا للمشاهدتين الأخيرتين لأغراض التنبؤ. باستخدام Microfit 4.1 حصلنا على قيم Y المتنبأ بها خلال 1997 و 1998 بالطريقة الاستاتيكية والديناميكية بناء على الانحدار المقدر خلال 1959 - 1996.

	Year 1997	Year 1998
Actual Y value	101.1	105.1
Static forecast of Y	107.24 (2.64)	109.45 (2.67)
Static forecast error	−6.14	−4.35
Dynamic forecast	100.75 (1.08)	101.95 (1.64)
Dynamic forecast error	0.35	3.14

لاحظ أن ما بين الأقواس يمثل الأخطاء القياسية المقدرة للقيم المتنبأ بها.

كما نرى من التمرين السابق، التنبؤات الديناميكية أقرب للقيم الفعلية من التنبؤات الاستاتيكية، والأخطاء القياسية للتنبؤات الديناميكية أقل من نظيرها الاستاتيكي. وبالتالي قد يكون من المفيد استخدام العملية AR(1) (أو عمليات أخرى برتب أعلى) لأغراض التنبؤ. عمومًا لاحظ أن كلاً من النوعين السابقين للتنبؤ فإن القيمة المتنبأ بها للأخطاء القياسية لـ 1998 أكبر من قيمة 1997، مما يعني أن التنبؤ بالمستقبل على فترات أبعد يشوبه بعض الخطر وهو أمر منطقي ومتوقع.

(47) لمزيد من التفاصيل، انظر Robert S. Pindyck and Daniel L. Rubinfeld, Econometric Models and Economic Forecasts, McGraw-Hill, 4th ed., 1998, pp. 214–217.

13.12 جوانب إضافية للارتباط الذاتي:
ADDITIONAL ASPECTS OF AUTOCORRELATION

المتغيرات الوهمية والارتباط الذاتي: Dummy Variables and Autocorrelation

في الفصل 9، استعرضنا نماذج انحدار المتغيرات الوهمية. وعلى وجه الخصوص نموذج انحدار .U.S والخاص بالدخل والادخار خلال الفترة 1970 - 1995 والذي قدمناه في (1.5.9) والذي يأخذ الشكل التالي:

$$Y_t = \alpha_1 + \alpha_2 + \beta_1 X_t + \beta_2 (D_t X_t) + u_t \qquad (1.13.12)$$

حيث Y = الادخار

X = الدخل

$D = 1$ للمشاهدات خلال 1982 - 1995

$D = 0$ للمشاهدات خلال 1970 - 1981

نتائج الانحدار وفقًا لهذا النموذج معطاة في (4.5.9). بالطبع هذا النموذج تم تقديره بناء على فروض OLS التقليدية.

والآن افترض أن u_t يتبع انحدارًا ذاتيًا من الدرجة الأولى، AR(1) أي أن $u_t = \rho u_{t-1} + \varepsilon_t$. مبدئيًا إذا كانت ρ معلومة أو من الممكن تقديرها باستخدام أي من الطرق السابق مناقشتها، فإنه من الممكن أن تستخدم طريقة الفروق العامة لتقدير معالم هذا النموذج والخالية من الارتباط الذاتي (من الدرجة الأولى). عمومًا وجود المتغير الوهمي D يفرض مشكلة خاصة وهي: لاحظ أن المتغير الوهمي ببساطة لقيم المشاهدات إما تتبع المرحلة الأولى أو المرحلة الثانية. كيف يمكننا تحويل ذلك؟ يمكن أن نتبع الأساليب التالية:[48]

1 - في (1.13.12). قيم D تساوي صفر لكل المشاهدات في الفترة الأولى، وفي الفترة الثانية، فإن قيمة D بالنسبة للمشاهدة الأولى تساوي $1/1-\rho$ بدلاً من 1 وتساوي 1 لباقي المشاهدات.

2 - المتغير X_t تم تحويله إلى $X_t - \rho X_{t-1}$. لاحظ أننا فقدنا مشاهدة واحدة في هذه التحويلة. إلا إذا استخدمنا تحويله Prais-Winsten واحتفظنا بالمفردة الأولى، كما سبق وذكرنا.

(48) Maddala, op. cit., pp. 321–322

3 - قيمة D_tX_t تساوي الصفر لكل المشاهدات الموجودة في الفترة الأولى (لاحظ : D_t تساوي الصفر في الفترة الأولى)، في الفترة الثانية، المشاهدة الأولى تأخذ القيمة $Y_t = D_tX_t$ وباقي المشاهدات في الفترة الثانية تساوي $(D_tX_t - D_t\rho X_{t-1}) = (X_t - \rho X_{t-1})$. (لاحظ أن قيمة D_t في الفترة الثانية تساوي 1).

كما يتضح من المناقشة السابقة، المشاهدة الحرجة هي المشاهدة الأولى في الفترة الثانية. إذا أخذنا ذلك في الاعتبار بالطريقة السابق ذكرها، فإنه من المفترض ألا توجد أي مشاكل في تقدير انحدار مثل (1.13.12) والمشتمل على ارتباط ذاتي AR(1). في تمرين 37.12، نسأل القارئ أن يقوم ببعض التحويلات على بيانات الدخل والإنفاق الخاصة بالـ U.S. والمعطاة في الفصل 9.

نماذج ARCH و GARCH :

كما أن حد الخطأ u عند الزمن t قد يكون مرتبطًا مع حد الخطأ عند الزمن $(t - 1)$ كما في العملية AR(1) أو بأي عدد من حدود الخطأ في فترات زمنية سابقة AR(p)، هل يمكن أيضًا أن يوجد ارتباط ذاتي للتباين σ^2 عند الزمن t مع قيمته في فترة زمنية سابقة واحدة أو أكثر؟ مثل هذا الارتباط الذاتي تم تواجده في العديد من الأبحاث خاصة المتعلقة بالسلاسل الزمنية المتنبأ بها في المجال المالي مثل أسعار الأسهم، معدلات التضخم، ومعدلات تغيير العملات. مثل هذا الارتباط الذاتي يأخذ تسمية أخرى، وهي ارتباط ذاتي مشروط باختلاف التباين (ARCH) في حالة ما إذا كان تباين الخطأ مرتبطًا بمربع حد الخطأ في الفترة السابقة، ويسمى الارتباط الذاتي العام المشروط باختلاف التباين (GARCH) إذا كان تباين الخطأ مرتبطًا بمربع حد الخطأ في عدد من الفترات الزمنية السابقة. وحيث إن هذا الموضوع يندرج أكثر تحت موضوعات السلاسل الزمنية في الاقتصاد القياسي، سنناقشه بعمق وبمزيد من التفصيل في الفصول المرتبطة بالسلاسل الزمنية في الاقتصاد القياسي.

الهدف هنا هو الإشارة إلى أن الارتباط الذاتي لا يقتصر على العلاقة بين حدود الخطأ الحالية والموجودة في فترات زمنية سابقة، ولكن أيضًا على تباينات الأخطاء الحالية والموجودة في فترات زمنية سابقة.

التواجد المشترك للارتباط الذاتي واختلاف التباين:
Coexistence of Autocorrelation and Heteroscedasticity

ماذا يحدث إذا كان نموذج الانحدار يعاني من كل من اختلاف التباين والارتباط الذاتي معًا؟ هل يمكن أن نحل المشكلة بشكل تتابعي، هل نبدأ أولاً باختلاف التباين، ثم نبحث بعد ذلك الارتباط الذاتي؟ في واقع الأمر أحد الكتاب قال التالي «الانحدار الذاتي يمكن اكتشافه فقط بعد أن يتم التحكم في اختلاف التباين». [49] ولكن هل يوجد اختبار ما يمكنه حل هذه المشكلة والمشاكل الأخرى (مثل توصيف النموذج) بشكل آنى؟ نعم، مثل هذه الاختبارات موجودة، ولكن مناقشتها تأخذنا بعيدًا عن نطاق هذا الكتاب. ومن الأفضل أن نترك هذه المواضيع لمن يهتم بها في قائمة المراجعة الخاصة بهذا الكتاب. [50]

14.12 الخلاصة والاستنتاجات: SUMMARY AND CONCLUSIONS

1 - إذا رفضنا الفرض الخاص بنموذج الانحدار الخطي التقليدي - القائل بأن حد الخطأ u_t والخاص بدالة انحدار المجتمع (PRF) عشوائي أو غير مرتبط، فإن مشكلة الارتباط الذاتي أو التسلسلي قد ظهرت في التحليل محل الدراسة.

2 - يظهر الارتباط الذاتي في البيانات لأسباب عديدة، مثل الطبيعة الخاملة للسلاسل الزمنية الاقتصادية، تحيز التوصيف نتيجة استبعاد متغيرات مهمة من النموذج أو استخدام شكل دالي غير سليم، ظاهرة بيت العنكبوت، تمسيد البيانات وتحويل البيانات. ومن المهم التفرقة بين الارتباط الذاتي المحض والارتباط الذاتي الراجع لواحد أو أكثر من الأسباب السابق ذكرها.

3 - على الرغم من وجود الارتباط الذاتي، إلا أن مقدرات OLS تظل غير متحيزة، متسقة وتؤول تقاربيًا إلى التوزيع الطبيعي، ولكن لم تعد هذه المقدرات كفئًا كما كانت. وكنتيجة لذلك، فإن اختبارات t، F و χ^2 لا يمكن تطبيقها مباشرةً، ويجب أن نحاول إصلاح ذلك.

(49) Lois W. Sayrs, Pooled Time Series Analysis, Sage Publications, California, 1989, p. 19.

(50) انظر Jeffrey M. Wooldridge, op. cit., pp. 402–403, and A. K. Bera and C. M. Jarque, "Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals: Monte Carlo Evidence," Economic Letters, vol. 7, 1981, pp. 313–318.

4 - طرق العلاج أو الإصلاح تعتمد على طبيعة العلاقة التبادلية الموجودة في حد الخطأ u_t. ولكن نظرًا لأن الأخطاء غير مشاهدة، فإن الطريقة الأشهر هي افتراض توليدها وفقًا لآلية ما.

5 - الآلية الأشهر تفترض عملية انحدار ذاتي لـ Markov من الدرجة الأولى، والتي تعنى أن الخطأ في الفترة الحالية مرتبط خطيًا مع الخطأ في الفترة الزمنية السابقة الأولى، معامل الارتباط الذاتي p يعبر عن مدى هذا الارتباط. هذه الآلية معروفة باسم عملية AR(1).

6 - إذا كانت عملية AR(1) صالحة للاستخدام، ومعامل الارتباط الذاتي معلومًا، فإن مشكلة الارتباط التسلسلي يمكن التغلب عليها مباشرةً عن طريق تحويل البيانات وفقًا لعملية الفروق العامة. يمكن بسهولة تعميم عملية AR(1) إلى AR(p)، ومن الممكن أيضًا افتراض آلية المتوسطات المتحركة MA أو خليط من AR و MA والمعروفة بـ ARMA. هذا الموضوع ستتم مناقشته في الفصول الخاصة بالسلاسل الزمنية في الاقتصاد القياسي.

7 - وإذا استخدمنا العملية AR(1) وكان معامل الارتباط الذاتي غير معلوم مسبقًا. هناك العديد من الطرق لتقدير ρ، مثل Durbin-Watson d المعدلة لـ Theil-Nagar، العملية التكرارية لـ Cochrane-Orcutt (C–O)، طريقة الخطوتين لـ C-O وطريقة الخطوتين لـ Durbin. في العينات كبيرة الحجم، هذه الطرق عادة ما تؤدي إلى نتائج متساوية لتقدير ρ، ولكنها في العينات صغيرة الحجم تختلف هذه التقديرات بشكل ملحوظ. من الناحية العملية، أخذت الطريقة التكرارية لـ C-O شهرة واسعة عند التطبيق.

8 - وفقًا لأي من الطرق السابق ذكرها، يمكننا استخدام طريقة الفروق العامة لتقدير معالم النموذج المحول باستخدام OLS، والتي تعتبر مساوية في هذه الحالة للـ GLS. ولكن نظرًا لأننا نقدر ρ $(\hat{\rho} =)$ نسمى طريقة التقدير طريقة ممكنة أو طريقة GLS المقدرة أو FGLS أو EGLS للاختصار.

9 - عند استخدام EGLS يجب على الباحث أن يضع في الاعتبار إسقاطه للمشاهدة الأولى من البيانات، في العينات صغيرة الحجم الاحتفاظ أو عدم الاحتفاظ

بالمشاهدة الأولى يمكن أن يكون له تأثير كبير على النتائج. وبالتالي في العينات صغيرة الحجم، يكون من الأفضل عمل تحويل للمشاهدة الأولى وفقًا لطريقة Prais-Winsten بالمشاهدة الأولى لا يكون له تأثير كبير على النتائج.

10 - من الجدير بالذكر، ملاحظة أن طريقة EGLE لها نفس الخصائص المثلى الإحصائية فقط في حالة العينات كبيرة الحجم، أما في الأحجام الصغيرة، OLS قد تكون أفضل من EGLS خصوصًا إذا كانت $\rho < 0 - 3$.

11 - بدلاً من استخدام EGLS، من الممكن استخدام OLS ولكن مع تصحيح الأخطاء القياسية للارتباط الذاتي وفقًا لطريقة Newey-West HAC. ولكن هذه الطريقة يمكن استخدامها فقط في أحجام العينات الكبيرة. إحدى مميزات استخدام طريقة HAC أنها لا تعالج مشكلة الارتباط الذاتي فقط، وإنما تعالج مشكلة اختلاف التباين إذا وجدت في البيانات محل الدراسة.

12 - بالطبع، قبل العلاج لابد أن يتم اكتشاف وجود ارتباط ذاتي في البيانات. هناك طرق رسمية وأخرى غير رسمية لاكتشاف الارتباط الذاتي. من بين الطرق أو الأساليب غير الرسمية، يمكن للباحث أن يرسم البواقي القياسية أو الحقيقية أو برسم البواقي الحالية ضد البواقي في فترات زمنية سابقة. ومن بين الطرق أو الأساليب الرسمية، يمكن استخدام اختبار الدفعات، اختبار Durbin-Watson d، اختبار الاعتيادية التقاربي، اختبار Berenblutt-Webb واختبار (BG) Breusch-Godfrey.

من كل هذه الطرق، فإن العديد من أحزمة الحاسب الآلي تستخدم اختبار d لـ Durbin-Watson. وعلى الرغم من تاريخ هذا الاختبار الطويل، إلا أن عليه قيود كبيرة. من الأفضل استخدام اختبار BG حيث إنه أكثر عمومية ويسمح بوجود الخطأ على الشكل AR أو MA كما يسمح بوجود المتغير المنحدر عليه في فترات زمنية سابقة كواحد من المتغيرات المفسرة.

ولكن ضع أيضًا في الاعتبار أنه اختبار صالح فقط للعينات ذات الأحجام الكبيرة.

13 - في هذا الفصل، ناقشنا أيضًا باختصار اكتشاف الارتباط الذاتي في حالة المتغيرات المنحدرة الوهمية، استخدام الأخطاء المرتبطة ذاتيًا في أغراض التنبؤ وموضوعات ARCH و GARCH.

تمـــارين : EXERCISES

أسئلة Questions

1.12 صح أم خطأ. مع التعليل

(a) عندما يوجد ارتباط ذاتي في البيانات، فإن مقدرات OLS متحيزة وغير كفء.

(b) اختبار Durbin-Watson d يفترض أن تباين حد الخطأ u_t ثابت.

(c) تحويلة الفروق الأولى والتي تعالج الارتباط الذاتي تفترض أن معامل الارتباط الذاتي ρ تساوي 1−.

(d) قيم R^2 لنموذج انحدار، واحد منهما في شكل الفروق الأولى، والآخر في المستوى الأساسي لا يمكن مقارنتهما ببعض مباشرةً.

(e) قيمة Durbin-Watson d المعنوية لا تعني بالضرورة وجود ارتباط ذاتي من الدرجة الأولى.

(f) في حالة وجود ارتباط ذاتي. القيم المحسوبة للتباينات والأخطاء القياسية للقيم المتنبأ بها غير كفء.

(g) عدم إدخال متغيرات مهمة في نموذج الانحدار قد يعطي قيمة لـ d معنوية.

(h) في عملية AR(1)، اختبار الفرض القائل بأن $\rho = 1$ يمكن التحقق منه من خلال إحصاء g لـ Berenblutt-Webb وأيضًا من خلال إحصاء d لـ Durbin-Watson.

(i) في انحدار الفروق الأولى لـ Y على الفروق الأولى لـ X، إذا كان هناك جزء ثابت، واتجاه عام خطي، فإن ذلك يعني أن النموذج الأصلي يشتمل على حد اتجاه عام خطي وتربيعي أيضًا.

2.12 إذا كان لدينا عينة مكونة من 50 مشاهدة و4 متغيرات مفسرة.. ما رأيك في وجود أو عدم وجود ارتباط ذاتي في البيانات إذا كان (a) $d = 1.05$؟ (b) $d = 1.40$؟ (c) $d = 2.50$؟ (d) $d = 3.97$؟.

3.12 في دراسة عن نصيب العاملين من الإنتاج وفقًا للقيمة المضافة (أي نصيب قوة العمل)، تم استخدام النماذج التالية وفقًا لـ Gujarati(*)

(*) Damodar Gujarati, "Labor's Share in Manufacturing Industries," Industrial and Labor Relations Review, vol. 23, no. 1, October 1969, pp. 65–75.

نموذج A : $Y_t = \beta_0 + \beta_1 t + u_t$

نموذج B : $Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + u_t$

حيث Y = نصيب العمالة، و t = الزمن. وفقًا لبيانات سنوية من الفترة 1949–1964. النتائج التالية تم الحصول عليها من الصناعات المعدنية الأساسية:

نموذج A : $\hat{Y}_t = 0.4529 - \underset{(-3.9608)}{0.0041t} \quad R^2 = 0.5284 \quad d = 0.8252$

نموذج B : $\hat{Y}_t = 0.4786 - \underset{(-3.2724)}{0.0127t} + \underset{(2.7777)}{0.0005t^2}$

$R^2 = 0.6629 \quad d = 1.82$

الأرقام بين الأقواس تمثل نسب t

(a) هل هناك ارتباط تسلسلي في النموذج A؟ النموذج B؟

(b) كيف علمت بوجود ارتباط تسلسلي؟

(c) كيف يمكنك التفرقة بين الارتباط الذاتي المحض، والارتباط الذاتي الرابع لتحيز التوصيف؟

4.12 اكتشاف الارتباط الذاتي. **اختبار النسب لـ von Neumann**(*). افترض أن البواقي $\hat{u}_t$ متغير عشوائي يتبع التوزيع الطبيعي. اثبت von Neumann أنه إذا كانت n كبيرة فإن النسبة

$$\frac{\delta^2}{s^2} = \frac{\sum(\hat{u}_i - \hat{u}_{i-1})^2/(n-1)}{\sum(\hat{u}_i - \bar{\hat{u}})^2/n}$$

لاحظ أن $\bar{\hat{u}} = 0$ في OLS

هذه النسبة تسمى نسبة von Neumann وهي تؤول تقاربيًا إلى التوزيع الطبيعي بتوقع

$$E\frac{\delta^2}{s^2} = \frac{2n}{n-1}$$

وتباين

$$\text{var}\,\frac{\delta^2}{s^2} = 4n^2\frac{n-2}{(n+1)(n-1)^3}$$

(a) إذا كانت n كبيرة بدرجة كافية. كيف يمكنك استخدام نسبة von Neumann لاختبار وجود ارتباط ذاتي في البيانات محل الدراسة؟

(*) J. von Neumann, "Distribution of the Ratio of the Mean Square Successive Difference to the Variance," Annals of Mathematical Statistics, vol. 12, 1941, pp. 367–395.

(b) ما هي العلاقة بين Durbin-Watson d وهذه النسبة؟

(c) احصاء d يقع بين 0 و 4. ما هي الحدود الخاصة بنسبة von Neumann؟

(d) بما أن النسبة تعتمد على فرض أن $\hat{u}$'s تتبع التوزيع الطبيعي. ما مدى صحة هذا الفرض بالنسبة لبواقي OLS؟

(e) افترض أنه عند التطبيق وجد أن النسبة تساوي 2.88 وكان حجم العينة 100 مفردة. اختر الفرض القائل بعدم وجود ارتباط تسلسلي في البيانات.

لاحظ أن: B.I.Hart قام بعمل جداول للقيم الحرجة لنسبة von Neumann للعينات حتى حجم 60 مفردة. (*)

5.12 في تسلسل من 17 من البواقي، 11 لها اشارة موجبة و 6 لها إشارة سالبة، عدد الدفعات يساوي 3. هل هناك دليل على وجود ارتباط ذاتي؟ هل ستتغير إجابتك إذا كان لدينا 14 دفعة؟

6.12 **تقدير ρ لـ Theil-Nagar بناء على الإحصاء d.** اقترح Theil-Nagar أنه في حالة العينات صغيرة الحجم بدلاً من تقدير ρ بـ $(1-\frac{d}{2})$. يتم تقديرها كالتالي:

$$\hat{\rho} = \frac{n^2(1-d/2)+k^2}{n^2-k^2}$$

حيث n = العدد الكلي للمشاهدات، d = Durbin-Watson d، و k = عدد المعاملات (مشتملة على الجزء الثابت المقطوع من المحور الصادي) المطلوب تقديرها.

اثبت أنه عندما تكون n كبيرة، تقدير ρ يتساوي مع التقدير الذي نحصل عليه من استخدام المعادلة الأبسط $(1-\frac{d}{2})$.

7.12 تقدير ρ : **عملية البحث لـ Hildreth-Lu**. (†) بما أن عملية الارتباط الذاتي من الدرجة الأولى تمثل كالتالي:

$$u_t = \rho u_{t-1} + \varepsilon_t$$

ρ من المتوقع أن تقع بين 1− و1+. اقترح Hildreth و Lu عملية بحث منظم لتحديد ρ. اقترحا اختبار ρ بين 1− و1+، مثلاً، 0.1 وحدة فترة وتحويل البيانات باستخدام معادلة الفروق العامة (5.6.12) وبالتالي يمكن اختبار ρ من 0.9−، 0.8−، . . . ، 0.8، 0.9 لكل قيمة مختارة من ρ تقوم بإجراء معادلة الفروق

(*) الجدول موجود في Johnston, op. cit., 3d ed., p. 559 .

(†) G. Hildreth and J. Y. Lu, "Demand Relations with Autocorrelated Disturbances," Michigan State University, Agricultural Experiment Station, Tech. Bull. 276, November 1960.

العامة، ونحصل على RSS المرافقة لها: $\Sigma \hat{u}_t^2$. اقترح Hildreth و Lu اختبار ρ التي تصغر قيمة RSS (أي تعظم قيمة R^2). إذا احتجنا تعديلاً، فإنه من الممكن تصغير وحدة الفترة أكثر، مثلاً 0.01 مثل 0.99–، 0.98–، . . . ، 0.90، 0.91 وهكذا.

(a) ما هي مميزات طريقة Hildreth-Lu؟

(b) كيف يمكن معرفة ما إذا كانت قيمة ρ المختارة لتحويل البيانات، ستؤدي في الحقيقة إلى تصغير $\Sigma \hat{u}_t^2$ ؟

8.12 تقدير ρ : العملية التكرارية لـ Cochrame-orcutt(*)

لتوضيح هذه العملية. اعتبر النموذج الثنائي المتغيرات التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (1)$$

والعملية AR(1)

$$u_t = \rho u_{t-1} + \varepsilon_t, -1 < \rho < 1 \quad (2)$$

أوصى Cochrane و Orcutt بإيقاع الخطوات التالية لتقدير ρ :

1 – قدر (1) باستخدام OLS العادية وأصل على البواقي $\hat{u}_t$ ، لاحظ أنه يمكن أن يكون لديك أكثر من متغير X واحد في النموذج.

2 – باستخدام البواقي التي حصلنا عليها في الخطوة 1. قم بعمل الانحدار التالي:

$$\hat{u}_t = \hat{\rho} \hat{u}_{t-1} + v_t \quad (3)$$

والذي يمثل النظير التطبيقي لـ (2). (†)

3 – باستخدام $\hat{\rho}$ التي حصلنا عليها في (3)، قدر معادلة الفروق العامة (6.9.12).

4 – بما أن القيمة غير معلومة مسبقًا. إذا كانت $\hat{\rho}$ التي حصلنا عليها من (3) تعتبر أفضل تقدير لـ ρ، استبدل قيم $\hat{\beta}_1^*$ و $\hat{\beta}_2^*$ التي حصلنا عليهما من (3) في النموذج الأصلي (1) واحصل على بواق جديدة مثلاً، $\hat{u}_t^*$ كالتالي:

(*) D. Cochrane and G. H. Orcutt, "Applications of Least-Squares Regressions to Relationships Containing Autocorrelated Error Terms," Journal of the American Statistical Association, vol. 44, 1949, pp 31–61.

(†) لاحظ أن $\hat{\rho} = \Sigma \hat{u}_t \hat{u}_{t-1} / \Sigma \hat{u}_t^2$ (لماذا؟) . على الرغم من التحيز ، $\hat{\rho}$ مقدر متسق لـ ρ الحقيقية .

$$\hat{u}_t^* = Y_t - \hat{\beta}_1^* - \hat{\beta}_2^* X_t \qquad (4)$$

والتي يمكن حسابها بسهولة، حيث إن Y_t، X_t، $\hat{\beta}_1^*$ و $\hat{\beta}_2^*$ معلومان.

5 - الآن قدر الانحدار التالي:

$$\hat{u}_t^* = \hat{\rho}^* \hat{u}_{t-1}^* + w_t \qquad (5)$$

والمماثل لـ (3)، وبالتالي يمثل المرحلة الثانية من تقدير ρ. بما أننا لا نعرف ما إذا كانت المرحلة الثانية لتقدير ρ أفضل أم لا لتقدير قيمة ρ الحقيقية، ننتقل إلى المرحلة الثالثة للتقدير وهكذا. وهذا ما يجعل عملية C-O تسمى العملية التكرارية. ولكن إلى أي مدى يمكننا الاستمرار في ذلك؟ القاعدة العامة هي أن تتوقف عندما تكون العمليات التكرارية لتقدير ρ لا تختلف عن بعضها البعض كثيراً، مثلاً أقل من 0.01 أو 0.05.

في مثالنا عن الأجور - الإنتاجية، احتجنا إلى سبع عمليات تكرارية قبل أن نتوقف.

(a) وفقًا لأى حزمة إلكترونية من اختيارك. اثبت أن القيمة المقدرة لـ ρ تساوي تقريبًا 0.8919 للمعادلة (16.9.12) و 0.961 للمعادلة (17.9.12).

(b) هل قيمة rho التي حصلنا عليها من عملية C-O تضمن نهاية صغرى عامة أم فقط نهاية صغرى محلية؟

(c) اختياري: طبق طريقة C-O على النموذج الخطي اللوغاريتمي للأجور والإنتاجية المعطى في (2.5.12)، مع الاحتفاظ بالمشاهدة الأولى مرة وعدم الاحتفاظ بها مرة أخرى. قارن نتائجك مع تلك التي حصلنا عليها من انحدار (1.5.12).

9.12 تقدير ρ: عملية الخطوتين لـ Cochrane-Orcutt. هذه تعتبر صيغة مختصرة للعملية التكرارية لـ C-O. في الخطوة 1، تقدير ρ من التكرار الأول، أي من المعادلة (3) في التمرين السابق، وفي الخطوة 2 تستخدم تقدير ρ وتقوم بعمل معادلة الفروق العامة كما في المعادلة (4) من التمرين السابق. أحيانًا عند التطبيق، تعطي طريقة الخطوتين نتائج مشابهة إلى حد كبير من تلك التي نحصل عليه من العملية التكرارية لـ C-O.

طبق طريقة الخطوتين لـ C-O على الانحدار التوضيحي للأجور - الإنتاجية المعطى من قبل، وقارن نتائجك مع تلك التي حصلنا عليها من الطريقة التكرارية. ضع في اعتبارك المشاهدة الأولى عند التحويل.

10.12 تقـديـر ρ : طريقـة الخطـوتين لـ Durbin .(*) لشـرح هذه الطريقـة دعنا نكتب معادلة الفروق العامة (5.9.12) في الشكل التالي :

$$Y_t = \beta_1(1-\rho) + \beta_2 X_t - \beta_2 \rho X_{t-1} + \rho Y_{t-1} + \varepsilon_t \qquad (1)$$

اقترح Durbin الطريقـة التاليـة ثنائيـة الخطوات لتقـدير ρ . أولاً استـخـدم (1) كنمـوذج انحـدار مـتعـدد، وقم بعـمل انحـدار لـ Y_t على Xt و X_{t-1} و Y_{t-1} ثم استخدم القيمة المقدرة لمعامل الانحدار لـ Y_{t-1} ($\hat{\rho}$ =) كتقدير لـ ρ . ثانيًا، بعد الحصول على $\hat{\rho}$ استخدم هذه القيمة للحصول على معالم معادلة الفروق العامة (5.9.12) أو مكافئيها (6.9.12) .

(a) طبق طريقة الخطوتين لـ Durbin على مثال الأجور - الإنتاجية الذي سبق وناقشناه، وقارن نتائجك مع تلك التي حصلنا عليها من العملية التكرارية لـ Cochrane-Orcutt وطريقـة الخطوتين لـ C-O . وعلق على مـدى جـودة النتائج التي حصلت عليها .

(b) إذا اختبرت المعادلة (1) الموجودة أعلى، ستلاحظ أن معامل X_{t-1} ($= -\rho\beta_2$) يساوي 1- مضروبًا في حاصل ضرب معامل X_t (β_2) ومعامل Y_{t-1} (ρ) . كيف يمكنك اختبار ما إذا كان هذا المعامل مستوفيًا قيود العملية أم لا؟

11.12 في قياسي العائد من المعروض من الكهرباء . استخدم Nerlove البيانات المقطعية لـ 145 منشأة منافع خاصة في الولايات المتحدة في الفترة 1955 وقام بعمل انحدار للوغاريتم التكلفة الكلية على لوغاريتم الناتج، معدل الأجور، سعر رأس المال وسعر الوقود . وجد أن البواقي المقدرة من هذا الانحدار تعاني من ارتباط «تسلسلي»، كما يؤكد ذلك قيمة d لـ Durbin-Watson للبحث عن علاج، قام برسم البواقي المقدرة ضد اللوغاريتم الناتج واحصل على الشكل (11.12) .

(a) ما الذي يمكنك استنتاجه من الشكل (11.12)؟

(b) كيف يمكنك التخلص من الارتباط «التسلسلي» في هذا الموقف؟

12.12 بواقي الانحدار عند رسمها في مقابل الزمن أعطت شكل الانتشار الموجود في الشكل (12.12) البواقي المتطرقة تسمى قيم شاذة . القيمة الشاذة هي مشاهدة

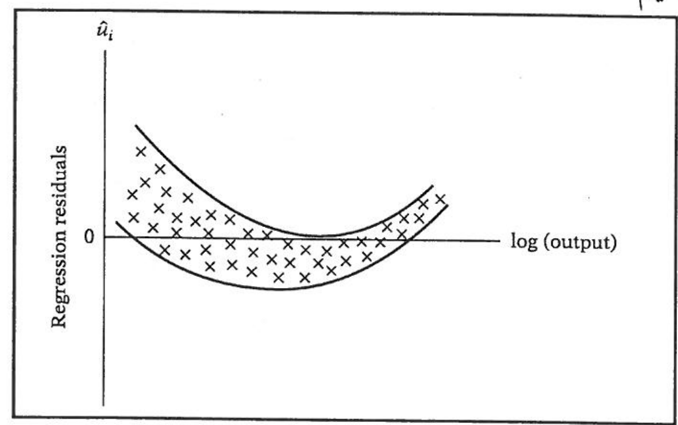
(*) J. Durbin, "Estimation of Parameters in Time-Series Regression Models," Journal of the Royal Statistical Society, Series B, vol. 22, 1960, p. 139–153

قيمتها تزيد عن قيم باقي المشاهدات في العينة بقيمة كبيرة، قد تكون 3 أو 4 انحراف معياري ± القيمة المتوقعة لكل المشاهدات.

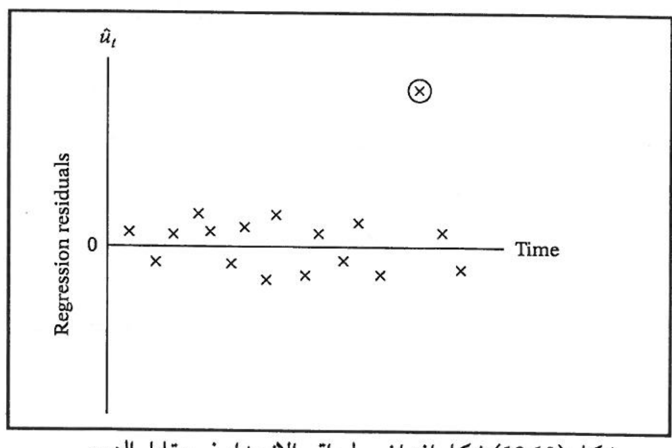
(a) ما هي أسباب وجود مثل هذه القيم الشاذة؟

(b) إذا وجدت قيمة شاذة، هل يجب حذف هذه المشاهدة وعمل الانحدار بناء على باقي المشاهدات فقط؟

(c) هل d لـ Durbin-Watson من الممكن تطبيقه في حالة وجود مفردة أو أكثر كقيم شاذة؟



شكل (11.12) بواقي الانحدار من دراسة Nerlove



شكل (12.12) شكل افتراضي لبواقي الانحدار في مقابل الزمن

13.12 بناء على إحصاء d لـ Durbin-Watson. كيف يمكنك التفرقة بين الارتباط الذاتي «المحض» وخطأ التحيز؟

14.12 افترض النموذج التالي:

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

u's في الحقيقة مستقلة. ماذا سيحدث في مثل هذا الموقف إذا افترضنا $u_t = \rho u_{t-1} + \varepsilon_t$، استخدم انحدار الفروق العام.

$$Y_t - \rho Y_{t-1} = \beta_1(1-\rho) + \beta_2 X_t - \rho\beta_2 X_{t-1} + \varepsilon_t$$

ناقش على وجه الخصوص خصائص حد الخطأ ε_t.

15.12 في دراسة لتحديد أسعار المنتج النهائي وفقًا للتكلفة في المملكة المتحدة، حصلنا على النتائج التالية بناء على بيانات سنوية خلال الفترة 1951 – 1969:

$$\widehat{PF}_t = 2.033 + 0.273W_t - 0.521X_t + 0.256M_t + 0.028M_{t-1} + 0.121PF_{t-1}$$

$$se = (0.992) \quad (0.127) \quad (0.099) \quad (0.024) \quad (0.039) \quad (0.119)$$

$$R^2 = 0.984 \qquad d = 2.54$$

حيث PF = سعر المنتج النهائي وفقًا للتكلفة، W = الأجور والرواتب للموظفين، X = الناتج المحلي الإجمالي لكل شخص موظف، M = أسعار الاستيراد، M_{t-1} = سعر الاستيراد في السنة السابقة و PF_{t-1} = سعر المنتج النهائي وفقًا للتكلفة في السنة السابقة. (*)

«بما أن لدينا 18 مشاهدة و 5 متغيرات مفسرة، فإن 5% حد أدنى وأعلى لقيمة d هي 0.71 و 2.06 بالترتيب، وقيمة d المقدرة تساوي 2.54. هذه القيمة تعني أنه لا يوجد ارتباط ذاتي طردي» علق على هذه العبارة.

16.12 استعرض الظروف التي تجعلنا نستخدم كل طريقة من الطرق التالية لتقدير معامل الارتباط الذاتي من الدرجة الأولى ρ:

(a) انحدار الفرق الأولى.

(b) انحدار المتوسطات المتحركة.

(c) تحويلة Theil-Nagar.

(d) العملية التكرارية لـ Orcutt و Cochrane.

(e) عملية الكشف لـ Hildreth-Lu.

(f) عملية الخطوتين لـ Durbin.

(*) المصدر: Prices and Earnings in 1951–1969: An Econometric Assessment, Department of Employment, Her Majesty's Stationery Office, 1971, Table C, p. 37, Eq. 63.

17.12 اعتبر النموذج التالي :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

حيث

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-1} + \varepsilon_t$$

أي أن حد الخطأ يتبع العملية AR(2) و ε_t حد خطأ عشوائي يتبع الفروض التقليدية . وضح الخطوات اللازم اتباعها لتقدير هذا النموذج، واضعًا في الاعتبار الانحدار الذاتي من الدرجة الثانية .

18.12 معادلة $\hat{\beta}_2^{\text{GLS}}$ المعطاه في (1.3.12) إذا تضمنت معامل التحديد C تصبح كالتالي :

$$\hat{\beta}_2^{\text{GLS}} = \frac{(1-\rho^2)x_1 y_1 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})(y_t - \rho y_{t-1})}{(1-\rho^2)x_1^2 + \sum_{t=2}^{n}(x_t - \rho x_{t-1})^2}$$

وفقًا لهذه المعادلة ولـ (1.3.12) . اوجد صيغة معامل التصحيح C .

19.12 اثبت أن تقدير (5.9.12) مساو لتقدير الـ GLS الذي سبق وناقشناه في الفقرة 3.12، أستبعد المشاهدة الأولى لـ Y و X .

20.12 القيمة المقدرة لبواقي انحدار (9.9.12) لها الإشارات التالية، وللتسهيل، هذه الإشارات تم وضعها بين أقواس

(++++)(−)(+++++++)(−)(++++)(−−)(+)(−−)(+)(−−)(++)(−)
(+)(−−−−−−−−−)(+)

وفقًا لاختبار الدفعات، هل ترفض الفرض العدمي القائل بأنه لا يوجد ارتباط ذاتي في البواقي؟

21.12(*) اختبار ارتباط تسلسلي من رتب عليا . افترض أن لدينا بيانات سلاسل زمنية على أساس ربع سنوي . في نماذج الانحدار الخاصة ببيانات ربع سنوية بدلاً من استخدام عملية AR(1) المعطاة في (1.2.12)، يكون من الأفضل استخدام AR(4) كالتالي :

$$u_t = \rho_4 u_{t-4} + \varepsilon_t$$

أي أننا نفترض أن حد الخطأ في الربع الحالي مرتبط بالخطأ الموجود في نفس ربع السنة محل الدراسة في السنة السابقة أكثر من ارتباطه بالخطأ في الربع التالي له مباشرةً .

(*) اختياري .

لاختبار الفرض القائل بأن $\rho_4 = 0$، اقترح Wallis(*) اختبار لـ d Durbin-Watson المعدل التالي:

$$d_4 = \frac{\sum_{t=5}^{n}(\hat{u}_t - \hat{u}_{t-4})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

الاختبار يتبع اختبار d التقليدي الذي سبق مناقشته. Wallis قدم جداول لـ d_4 ويمكن للقارئ أن يطلع عليها في البحث الأصلي لـ Wallis.

افترض الآن أن لدينا بيانات شهرية. هل يمكن تعميم اختبار Durbin-Watson ليشمل هذه الحالة؟ إذا كان ذلك ممكنًا اكتب صيغة d_{12} المناسبة.

22.12 افترض أننا نقدر النموذج التالي:

$$\Delta \ln \text{output}_t = \beta_1 + \beta_2 \Delta \ln L_t + \beta_3 \Delta \ln K_t + u_t$$

حيث Y تمثل الناتج، L العمالة، K رأس المال و Δ تمثل معامل الفروق الأولى. كيف يمكنك تفسير β_1 في هذا النموذج؟ هل يمكننا اعتبارها تقديراً للتغير التكنولوجي؟ علل إجابتك.

23.12 كما سبق وذكرنا، فقد اقترح Maddala أنه إذا كانت قيمة d لـ Durbin-Watson أصغر من R^2، يفضل عمل الانحدار في شكل الفروق الأولى. ما هو تفسيرك لهذا الاقتراح؟

24.12 بالعودة إلى المعادلة (1.4.12)، افترض أن $r = 0$ ولكن $\rho \neq 0$. ماهو أثر ذلك على $E(\hat{\sigma}^2)$ إذا كانت (a) $0 < \rho < 1$ و(b) $-1 < \rho < 0$؟ متى سيكون التحيز الموجود في $\hat{\sigma}^2$ صغيراً بشكل مقبول؟

25.12 قمنا بعمل انحدار لبواقي نموذج الأجور - الإنتاجية المعطى في (1.5.12) على البواقي في فترات زمنية سابقة، ست فترات زمنية سابقة [أي AR(6)]. حصلنا على النتائج التالية:

(*) Kenneth Wallis, "Testing for Fourth Order Autocorrelation in Quarterly Regression Equations," Econometrica, vol. 40, 1972, pp. 617–636. Tables of d_4 can also be found in J. Johnston, op. cit., 3d ed., p. 558.

```
Dependent Variable: RES1
Method: Least Squares
Sample(adjusted): 1965-1998
Included Observations: 34 after adjusting endpoints
```

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5.590462	1.963603	2.847043	0.0085
X	-0.066605	0.023469	-2.838058	0.0087
RES1(-1)	0.814971	0.216231	3.768978	0.0009
RES1(-2)	-0.268651	0.273887	-0.980882	0.3357
RES1(-3)	-0.106017	0.272780	-0.388652	0.7007
RES1(-4)	0.305630	0.273258	1.118467	0.2736
RES1(-5)	-0.064375	0.280577	-0.229438	0.8203
RES1(-6)	0.216156	0.222160	0.972976	0.3395

$R^2 = 0.8920$ Durbin-Watson d stat 1.7589
$\bar{R}^2 = 0.8629$

(a) من النتائج السابقة، ما الذي تستنتجه عن طبيعة الارتباط الذاتي في بيانات الأجور - الإنتاجية؟

(b) إذا اعتقدت أنه يمكن استخدام العملية AR(1) للتعبير عن الارتباط الذاتي في البيانات، هل ستستخدم تحويلة الفروق الأولى للتخلص من الارتباط الذاتي؟ علل إجابتك.

Problems مسائل:

26.12 بالعودة إلى بيانات صناعة النحاس المعطاة في جدول (7.12)

(a) بناء على هذه البيانات، قدر نموذج الانحدار التالي:

$$\ln C_t = \beta_1 + \beta_2 \ln I_t + \beta_3 \ln L_t + \beta_4 \ln H_t + \beta_5 \ln A_t + u_t$$

فسر نتائجك.

(b) احصل على البواقي والبواقي القياسية من النموذج السابق وارسمها ما الذي يمكنك استنتاجه عن طبيعة الارتباط الذاتي في هذه البواقي؟

(c) قدر قيمة إحصاء d لـ Durbin-Watson وعلق على طبيعة الارتباط الذاتي الموجود في البيانات.

(d) قم عمل اختبار الدفعات، وهل نتائجك ستختلف عن تلك التي حصلت عليها في C؟

(e) كيف يمكنك معرفة ما إذا كانت العملية AR(p) توصف بالارتباط الذاتي الموجود في البيانات محل الدراسة أفضل من العملية AR(1)؟

لاحظ التالي : احتفظ بالبيانات ، حيث سيتم استخدامها مرة أخرى في تحاليل أخرى تالية (انظر تمرين 28.12).

27.12 وفقًا للبيانات المعطاة في جدول (8.12).

(a) اثبت أن قيمة d لـ Durbin-Watson = 0.41148 .

(b) هل هناك ارتباط تسلسلي طردي في الخطأ؟

جدول (7.12) محددات السعر المحلي للنحاس في الولايات المتحدة خلال 1951 – 1980

Year	C	G	I	L	H	A
1951	21.89	330.2	45.1	220.4	1,491.0	19.00
52	22.29	347.2	50.9	259.5	1,504.0	19.41
53	19.63	366.1	53.3	256.3	1,438.0	20.93
54	22.85	366.3	53.6	249.3	1,551.0	21.78
55	33.77	399.3	54.6	352.3	1,646.0	23.68
56	39.18	420.7	61.1	329.1	1,349.0	26.01
57	30.58	442.0	61.9	219.6	1,224.0	27.52
58	26.30	447.0	57.9	234.8	1,382.0	26.89
59	30.70	483.0	64.8	237.4	1,553.7	26.85
60	32.10	506.0	66.2	245.8	1,296.1	27.23
61	30.00	523.3	66.7	229.2	1,365.0	25.46
62	30.80	563.8	72.2	233.9	1,492.5	23.88
63	30.80	594.7	76.5	234.2	1,634.9	22.62
64	32.60	635.7	81.7	347.0	1,561.0	23.72
65	35.40	688.1	89.8	468.1	1,509.7	24.50
66	36.60	753.0	97.8	555.0	1,195.8	24.50
67	38.60	796.3	100.0	418.0	1,321.9	24.98
68	42.20	868.5	106.3	525.2	1,545.4	25.58
69	47.90	935.5	111.1	620.7	1,499.5	27.18
70	58.20	982.4	107.8	588.6	1,469.0	28.72
71	52.00	1,063.4	109.6	444.4	2,084.5	29.00
72	51.20	1,171.1	119.7	427.8	2,378.5	26.67
73	59.50	1,306.6	129.8	727.1	2,057.5	25.33
74	77.30	1,412.9	129.3	877.6	1,352.5	34.06
75	64.20	1,528.8	117.8	556.6	1,171.4	39.79
76	69.60	1,700.1	129.8	780.6	1,547.6	44.49
77	66.80	1,887.2	137.1	750.7	1,989.8	51.23
78	66.50	2,127.6	145.2	709.8	2,023.3	54.42
79	98.30	2,628.8	152.5	935.7	1,749.2	61.01
80	101.40	2,633.1	147.1	940.9	1,298.5	70.87

لاحظ أن: البيانات جمعها Gray R. Smith من مصادر مختلفة مثل السوق الأمريكية للمعادن ، أسبوع المعادن ومنشورات القطاع التجاري الأمريكي .

C = متوسط سنوي السعر المحلي للنحاس في الولايات المتحدة (سنت لكل بوند)

G = الناتج المحلي الكلي السنوي (بليون دولار)

I = مؤشر الإنتاج الصناعي لـ 12 شهراً .

L = متوسط السعر التبادلي للمعادن في لندن (متوسط سنوي = 12 شهراً)

H = عدد المنازل في السنة (آلاف الوحدات)

A = متوسط سعر الالومنيوم (متوسط سنوي = 12 شهراً) (سنت لكل بوند)

(c) وفقًا لكل ما سبق، قدر ρ كالتالي:

i – طريقة Theil-Nagar

ii – عملية الخطوتين لـ Durbin

iii – طريقة Cochrane-Orcutt

(d) استخدم طريقة Theil-Nagar لتحويل البيانات، ثم قم بعمل الانحدار باستخدام البيانات المحولة.

(e) هل الانحدار المقدر في d يوجد قيمة ارتباط ذاتي؟ إذا حدث ذلك فعلاً، كيف يمكن التخلص منه؟

28.12 بالعودة إلى تمرين (26.12) وباستخدام البيانات الموجودة في جدول (7.12). وإذا أظهرت النتائج ارتباطًا تسلسليًا:

(a) استخدم عملية الخطوتين لـ Cochrane-Orcutt واصل على تقدير لـ GLS الممكنة أو انحدار الفروق العامة وقارن نتائجك.

(b) إذا كانت ρ المقدرة من طريقة Cochrane-Orcutt في ρ تختلف بشكل كبير عن القيمة المقدرة باستخدام إحصاء d، أي من الطريقتين ستختار لتقدير ρ ولماذا؟

جدول (8.12)

Y, personal consumption expenditure, billions of 1958 dollars	X, time	$\hat{Y}$, estimated Y*	$\hat{u}$, residuals
281.4	1 (= 1956)	261.4208	19.9791
288.1	2	276.6026	11.4973
290.0	3	291.7844	−1.7844
307.3	4	306.9661	0.3338
316.1	5	322.1479	−6.0479
322.5	6	337.3297	−14.8297
338.4	7	352.5115	−14.1115
353.3	8	367.6933	−14.3933
373.7	9	382.8751	−9.1751
397.7	10	398.0569	−0.3569
418.1	11	413.2386	4.8613
430.1	12	428.4206	1.6795
452.7	13	443.6022	9.0977
469.1	14	458.7840	10.3159
476.9	15 (= 1970)	473.9658	2.9341

* تم الحصول على هذا الجدول من الانحدار $Y_t = \beta_0 + \beta_1 X_t + u_t$

29.12 بالعودة إلى مثال 4.7. احذف المتغيرين X_2 و X_3 ثم قم بعمل الانحدار، واختبر الارتباط «التسلسلي» للبواقي. إذا وجدت أن هناك ارتباطًا تسلسليًا، كيف يمكنك تفسيره؟ ما هي الخطوات العلاجية التي يمكن أن تقترحها؟

30.12 بالعودة إلى تمرين 21.7. هناك ارتباط ذاتي مسبق متوقع في هذه البيانات. وبالتالي نقترح أن تقوم بعمل انحدار للوغاريتم المعروض الحقيقي من المال على لوغاريتم الدخل القومي الحقيقي ومعدل الفائدة طويل المدى في شكل الفروق الأولى. قم بعمل مثل هذا الانحدار، وقم بإعادة الانحدار مرة أخرى في الشكل الأصلي. هل الفرض الخاص بتحويلة الفروق الأولى مستوفى؟ إذا لم يكن مستوفيًا، ما هو شكل التحيز المفترض وجوده في النتائج بسبب استخدام هذه التحويلة؟ اشرح فكرتك باستخدام البيانات المعطاة.

31.12 **استخدام d لـ Durbin-Watson لاختبار عدم الخطية.**

بالعودة إلى تمرين 29.12. رتب البواقي التي حصلت عليها من الانحدار تصاعديًا وفقًا لقيم X. باستخدام المعادلة المعطاة في (5.6.12) قدر d من البواقي المرتبة. إذا كانت قيمة d المحسوبة تشير إلى وجود ارتباط ذاتي، هذا قد يعني أن النموذج الخطي غير سليم، وأن النموذج الصحيح قد يشتمل على X_i^2 أو X_i^3. هل يمكنك إعطاء تعليل قوى لذلك؟ تحقق مما إذا كانت إجابتك متماشية مع Henri Theil. (*)

32.12 بالعودة إلى تمرين 22.11. احصل على البواقي، وحدد ما إذا كانت تعاني من الارتباط الذاتي أم لا. كيف يمكنك تحويل البيانات في حالة وجود ارتباط ذاتي؟ ما هو معنى الارتباط التسلسلي في هذه الحالة؟

33.12 تجربة Monte Carlo. بالعودة إلى الجدولين (1.12) و(2.12) باستخدام بيانات ε_t و X_t المعطاة، ولدت عينة من 10 مشاهدات لـ Y وفقًا للنموذج التالي:

$$Y_t = 3.0 + 0.5\, Y_t + u_t$$

حيث $u_t = 0.94u_{t-1} + \varepsilon_t$ وافترض أن $u_0 = 10$.

(a) قدر المعادلة وعلق على النتائج.

(b) افترض الآن أن $u_0 = 17$. كرر التمرين عشر مرات، وعلق على النتائج.

(c) اتبع نفس الخطوات السابقة فيما عدا أن $\rho = 0.3$ بدلاً من $\rho = 0.9$ وقارن بين نتائجك الحالية ونظيرها الذي حصلت عليه في b.

(*) Henri Theil, Introduction to Econometrics, Prentice Hall, Englewood Cliffs, N. J., 1978, pp. 307–308.

34.12 باستخدام البيانات المعطاة في جدول (9.12)، قدر النموذج التالي :

$$Y_t = \beta_1 + \beta_2 X_t + u_t$$

حيث Y = المخزون و X = المبيعات، كلاهما مقاس بالبليون دولار.

جدول (9.12) المخزون والمبيعات في مصانع الولايات المتحدة خلال 1950 – 1991 (مليون دولار)

Inventories and Sales in U.S. Manufacturing 1950-1991 (Millions of Dollens)

Year	Sales*	Inventories†	Year	Sales*	Inventories†
1950	38,596	59,822	1970	108,352	178,594
1951	43,356	70,242	1971	117,023	188,991
1952	44,840	72,377	1972	131,227	203,227
1953	47,987	76,122	1973	153,881	234,406
1954	46,443	73,175	1974	178,201	287,144
1955	51,694	79,516	1975	182,412	288,992
1956	54,063	87,304	1976	204,386	318,345
1957	55,879	89,052	1977	229,786	350,706
1958	54,201	87,055	1978	260,755	400,929
1959	59,729	92,097	1979	298,328	452,636
1960	60,827	94,719	1980	328,112	510,124
1961	61,159	95,580	1981	356,909	547,169
1962	65,662	101,049	1982	348,771	575,486
1963	68,995	105,463	1983	370,501	591,858
1964	73,682	111,504	1984	411,427	651,527
1965	80,283	120,929	1985	423,940	665,837
1966	87,187	136,824	1986	431,786	664,654
1967	90,918	145,681	1987	459,107	711,745
1968	98,794	156,611	1988	496,334	767,387
1969	105,812	170,400	1989	522,344	813,018
			1990	540,788	835,985
			1991	533,838	828,184

المصدر: Economic Report of the President, 1993, Table B-53, p. 408

(*) البيانات السنوية عبارة عن متوسطات شهرية، غير معدلة موسميًا.

(†) بيانات معدلة موسميًا، الفترة النهائية التي تبدأ 1982 لا تقارن بالفترة السابقة.

(a) قدر الانحدار السابق.

(b) باستخدام البواقي المقدرة، هل يوجد ارتباط ذاتي طردي باستخدام (i) اختبار Durbin-Watson و(ii) اختبار الاعتيادية لأحجام العينات الكبيرة المعطى في (13.6.12).

(c) إذا كانت ρ موجبة، طبق اختبار Berenblutt-Webb لاختبار الفرض القائل بأن $\rho = 1$.

(d) إذا كان عندك شك أن نمط الارتباط الذاتي للأخطاء هو p، استخدم اختبار Breusch-Godfrey لإثبات ذلك. كيف تختار قيمة p؟

(e) وفقًا لنتائج هذا الاختبار، كيف يمكنك تحويل البيانات للتخلص من الارتباط الذاتي؟ وضح كل الخطوات الحسابية.

(f) كرر الخطوات السابقة وفقًا للنموذج التالي:

$$\ln Y_t = \beta_1 + \beta_2 \ln X_t + u_t$$

(g) كيف يمكنك أن تختار بين التوصيف الخطي أو الخطي اللوغاريتمي؟ وضح الاختبار المستخدم.

35.12 جدول (10.12) يعطي بيانات عن معدل العائد الحقيقي على بعض الاسم عند الزمن (RR_t)، معدل نمو الناتج في الفترة ($t+1$)، (OG_{t+1})، والتضخم عند الزمن t ($\ln f_t$)، كل البيانات في صورة نسب خاصة بالاقتصاد الأمريكي خلال الفترة 1954 إلى 1981.

(a) قم بعمل انحدار لـ RR_t على التضخم.

(b) قم بعمل انحدار لـ RR_t على OG_{t-1} و Inf_t.

(c) علق على الانحدارين السابقين وفقًا لملاحظة Eugene Fama والتي تنص على «الارتباط البسيط العكسي بين عائد الأسهم الحقيقي والتضخم ارتباطًا مزيفًا، فهو ليس إلا نتيجة لعلاقتين منظمتين: الأولى هي علاقة طردية بين عائد الأسهم الحقيقي الحالي والمعدل المتوقع لنمو الناتج [مقاس بـ OG_{t+1}] والثانية علاقة عكسية بين المعدل المتوقع لنمو الناتج والتضخم الحالي».

(d) هل تتوقع وجود ارتباط ذاتي في أي من الانحدارين السابقين **a** و **b**؟ علل إجابتك؟ وفقًا لإجابتك اتخذ الإجراءات المناسبة لعلاج ذلك واستعرض النتائج الصحيحة مرة أخرى.

جدول (10.12) معدل العائد، معدل نمو الناتج والتضخم، الولايات المتحدة 1965–1981

Observation	RR	Growth	Inflation
1954	53.0	6.7	−0.4
1955	31.2	2.1	0.4
1956	3.7	1.8	2.9
1957	−13.8	−0.4	3.0
1958	41.7	6.0	1.7
1959	10.5	2.1	1.5
1960	−1.3	2.6	1.8
1961	26.1	5.8	0.8
1962	−10.5	4.0	1.8
1963	21.2	5.3	1.6
1964	15.5	6.0	1.0
1965	10.2	6.0	2.3

تابع - جدول (10.12) معدل العائد ، معدل نمو الناتج والتضخم ، الولايات المتحدة 1965 - 1981 .

Observation	RR	Growth	Inflation
1966	−13.3	2.7	3.2
1967	21.3	4.6	2.7
1968	6.8	2.8	4.3
1969	−13.5	−0.2	5.0
1970	−0.4	3.4	4.4
1971	10.5	5.7	3.8
1972	15.4	5.8	3.6
1973	−22.6	−0.6	7.9
1974	−37.3	−1.2	10.8
1975	31.2	5.4	6.0
1976	19.1	5.5	4.7
1977	−13.1	5.0	5.9
1978	−1.3	2.8	7.9
1979	8.6	−0.3	9.8
1980	−22.2	2.6	10.2
1981	−12.2	−1.9	7.3

36.12 إحصاء h لـ Durbin : اعتبر نموذج محددات الأجر التالي :

$$Y_t = \beta_1 + \beta_2 X_t + \beta_3 Y_{t-1} + u_t$$

حيث Y = الأجور = مؤشر التعويض نصف الساعة .

X = الإنتاجية = مؤشر الناتج لكل ساعة .

(a) باستخدام بيانات جدول (4.12)، قدر النموذج السابق . مع تفسير النتائج .

(b) بما أن النموذج يشتمل على قيم المتغير المنحدر عليه في فترة زمنية سابقة كمتغير منحدر، فإن d لـ Durbin-Watson لا يجوز استخدامه لاكتشاف ما إذا كان هناك ارتباط ذاتي في البيانات أم لا . لمثل هذه النماذج والمسمى نماذج الانحدار الذاتي، قام Durbin بعمل إحصاء جديد يسمى إحصاء h لاختبار الارتباط الذاتي من الدرجة الأولى والمعرف كالتالي : (*)

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\text{var}\,(\hat{\beta}_3)]}}$$

حيث n = حجم العينة، $\text{var}\,(\hat{\beta}_3)$ = تباين معامل Y_{t-1} و $\hat{\rho}$ = مقدر الارتباط التسلسلي من الدرجة الأولى .

(*) J. Durbin, "Testing for Serial Correlation in Least-squares Regression When Some of the Regressors Are Lagged Dependent Variables," Econometrica, vol. 38, pp. 410–421.

للأحجام العينات الكبيرة (أو تقاربيًا)، اثبت Durbin أنه تحت صحة الفرض العدمي ($\rho = 0$) فإن

$$h \sim N(0\,,\,1)$$

أي أن إحصاء h يتبع التوزيع الطبيعي القياسي، من خصائص التوزيع الطبيعي نعلم ان احتمال $|h| > 1.96$ يساوي تقريبًا 5%. وبالتالي إذا كان في أي تطبيق $|h| > 1.96$، تستطيع رفض الفرض العدمي الخاص بأن $\rho = 0$، أي أنه يوجد دليل على وجود ارتباط ذاتي من الدرجة الأولى في نموذج الانحدار الذاتي المعطى أعلى.

لتطبيق الاختبار، ستقوم بعمل التالي: أولاً، قدر النموذج باستخدام OLS (لا تهتم بمشاكل التقدير عن المرحلة الحالية). ثانيًا، سجل قيمة $\text{var}\,(\hat{\beta}_3)$ في هذا النموذج بالإضافة إلى إحصاء d. ثالثًا، باستخدام قيمة d احصل على $\hat{\rho} \approx (1 - \frac{d}{2})$. ومن المثير للانتباه هنا أنه على الرغم من أننا لا نستطيع استخدام قيمة d لاختبار الارتباط التسلسلي في النموذج، إلا أننا نستخدمها للحصول على تقدير ρ. رابعًا نحسب الآن الإحصاء h. خامسًا، إذا كان حجم العينة كبيرًا بدرجة كافية وإذا زادت قيمة $|h|$ المحسوبة عن 1.96، نستنتج أن هناك دليلاً لوجود ارتباط ذاتي من الدرجة الأولى. بالطبع تستطيع استخدام أي مستوى آخر من المعنوية ترغب فيه.

طبق اختبار h على نموذج الانحدار الذاتي الخاص بمحددات الأجر والمعطى سابقًا، واستعرض نتائجك مع مقارنتها للنتائج المعطلة في انحدار (1.5.12).

37.12 المتغيرات الوهمية والارتباط الذاتي. بالعودة إلى انحدار الدخل والادخار السابق مناقشته في الفصل 9. استخدم البيانات المعطاة في جدول (2.9) وبافتراض العملية AR(1) أعد تقدير انحدار الدخل – الادخار ويأخذ في الاعتبار الارتباط الذاتي. اهتم بشكل خاص بتحويلة المتغير الوهمي. قارن نتائجك مع نظيرها المقدم في الفصل 9.

38.12 باستخدام بيانات الأجر – الإنتاجية المعطاة في جدول (4.12)، قدر النموذج (8.9.12) وقارن نتائجك مع نظيرها الذي حصلت عليه في (9.9.12). ما الذي تستنتجه؟

Appendix 12A ملحق A 12

1.A12 إثبات أن حد الخطأ v_t الموجود في (11.1.12) مرتبط ذاتيا:

Proof that the error term v_t in (12.1.11) is Autocorrelated

بما أن $v_t = u_t - u_{t-1}$، فمن السهل إثبات أن

$E(v_t) = E(u_t - u_{t-1}) = E(u_t) - E(u_{t-1}) = 0$، بما أن $E(u)=0$ لأي t.

الآن $\text{var}(v_t) = \text{var}(u_t - u_{t-1}) = \text{var}(u_t) + \text{var}(u_{t-1}) = 2\sigma^2$

بما أن تباين كل u_t يساوي σ^2 و u's مستقلة. إذن v_t له تباين ثابت ولكن

$$\begin{aligned}\text{cov}(v_t, v_{t-1}) &= E(v_t v_{t-1}) = E[(u_t - u_{t-1})(u_{t-1} - u_{t-2})] \\ &= -\sigma^2\end{aligned}$$

والذي لا يساوي الصفر. وبالتالي على الرغم من أن u's غير مرتبطة ذاتيًا، إلا أن v's مرتبطة ذاتيًا.

2.A12 إثبات المعادلات (3.2.12)، (4.2.12) و (5.2.12)

Proof of equations (12.2.3), (12.2.4) and (12.2.5)

وفقًا لـ AR(1)

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{1}$$

وبالتالي:

$$E(u_t) = \rho E(u_{t-1}) + E(\varepsilon_t) = 0 \tag{2}$$

و

$$\text{var}(u_t) = \rho^2 \,\text{var}(u_{t-1}) + \text{var}(\varepsilon_t) \tag{3}$$

حيث إن u's و ε's غير مرتبطين

بما أن $\text{var}(u_t) = \text{var}(u_{t-1}) = \sigma^2$ و $\text{var}(\varepsilon_t) = \sigma_\varepsilon^2$

فإننا نحصل على

$$\text{var}(u_t) = \frac{\sigma_\varepsilon^2}{1-\rho^2} \tag{4}$$

والآن بضرب (1) في u_{t-1} وإدخال التوقع على الطرفين نحصل على

$$\text{cov}(u_t, u_{t-1}) = E(u_t u_{t-1}) = E\left[\rho u_{t-1}^2 + u_{t-1}\varepsilon_t\right] = \rho E\left(u_{t-1}^2\right)$$

لاحظ أن التغايرين u_{t-1} و ε_t يساويان الصفر (لماذا؟)

و $$\text{var}(u_t) = \text{var}(u_{t-1}) = \frac{\sigma_\varepsilon^2}{1-\rho^2}$$

وبالتالي نحصل على

$$\text{cov}(u_t, u_{t-1}) = \rho\frac{\sigma_\varepsilon^2}{(1-\rho^2)} \qquad (5)$$

وبالاستمرار على نفس النهج نحصل على

$$\text{cov}(u_t, u_{t-2}) = \rho^2\frac{\sigma_\varepsilon^2}{(1-\rho^2)}$$

$$\text{cov}(u_t, u_{t-3}) = \rho^3\frac{\sigma_\varepsilon^2}{(1-\rho^2)}$$

وهكذا .

الآن معامل الارتباط هو نسبة التغاير إلى التباين وبالتالي :

$$\text{cor}(u_t, u_{t-1}) = \rho$$

$$\text{cov}(u_t, u_{t-2}) = \rho^2$$

وهكذا .

الفصل الثالث عشر

نمذجة الاقتصاد القياسي: توصيف النموذج واختبارات التشخيص

ECONOMETRIC MODELING: MODEL SPECIFICATION AND DIAGNOSTIC TESTING

لا يمكن تطبيق الاقتصاد القياسي بشكل آلي، فلابد من فهم تداعيات الموقف والمهارات المتاحة[1].

عندما نعبر كوبرى ما، لا يكون لدينا أي قلق من انهياره، نظرًا للثقة الشديدة في المهندسين الذين قاموا بتصميم وبناء هذا الكوبرى. على الاقتصاديين أن يفعلوا الشيء ذاته في النماذج الاقتصادية، وإلا عليهم أن يلصقوا التحذير التالي بالنماذج المختارة " لا تستخدمه فقد ينهار بك ! ". [2]

الاقتصاديون يبحثون عن "الحقيقة" على مر السنوات السابقة، على أساس أن الاقتصاديين أشخاص يبحثون عن قطة سوداء غير موجودة أصلاً في حجرة مظلمة، وبالتالي مستحيل أن يجدوها. [3]

أحد فروض نموذج الانحدار الخطي التقليدي (CLRM)، الفرض 9. إن نموذج الانحدار المستخدم في التحليل "صحيح" التوصيف، وإذا كان النموذج غير "صحيح" التوصيف، فإننا نتعرض إلى المشكلة المعروفة باسم "خطأ توصيف النموذج" أو تحيز توصيف النموذج. في هذا الفصل، سنلقي الضوء على مثل هذا الفرض، لأن البحث عن النموذج الصحيح يشبه البحث عن Grail المقدس.
بالتحديد سنناقش الأسئلة التالية:

---

(1) Keith Cuthbertson, Stephen G. Hall, and Mark P. Taylor, Applied Econometrics Techniques, Michigan University Press, 1992, p. X.

(2) David F. Hendry, Dynamic Econometrics, Oxford University Pree, U. K., 1995, p. 68.

(3) Peter Kennedy, A Guide to Econometrics, 3d ed., The MIT Press, Cambridge, Mass., 1992, p. 82.

1 - كيف يمكن للباحث تحديد النموذج "الصحيح" ؟ أو بمعنى آخر، ما هي الأدوات التي يمكن استخدامها لتحديد النموذج المناسب للتحليل التطبيقي؟

2 - ما هي أخطاء توصيف النموذج المحتملة عند التطبيق؟

3 - ما هي عواقب أخطاء التوصيف؟

4 - كيف يمكن للباحث أن يكتشف خطأ التوصيف؟ أو بمعنى آخر ما هي بعض أدوات التشخيص التي يمكن للباحث أن يستخدمها؟

5 - إذا كانت هناك أخطاء في التوصيف، ما هي الأساليب العلاجية الممكن استخدامها وما هي فوائدها؟

6 - كيف يمكن تقييم أداء النماذج المنافسة؟

موضوع توصيف النموذج وتقييمه يعتبر موضوعًا متعدد الأبعاد، وله العديد من التطبيقات. ليس ذلك فقط ولكن هناك فروقًا فلسفية وراء النماذج المختلفة في هذا الموضوع. بالطبع لن نستطيع أن نغطي كل ذلك في فصل واحد، ولكن نتمنى أن نلقى الضوء على بعض المواضيع الأساسية الخاصة بتوصيف النموذج وتقييمه.

1.13 معيار اختيار النموذج: MODEL SELECTION CRITERIA

بناء على دراسة Richard و Hendry فإن المعايير التي يجب استيفاؤها لاختيار نموذج ما للتحليل التطبيقي هي كالتالي: [4]

1 - **أن يكون في إطار البيانات المتاحة**: أي لابد أن تكون التنبؤات المتوقعة من النموذج يمكن فعليًا ومنطقيًا الحصول عليها.

2 - **الاتساق مع النظرية**: فلابد أن يكون هناك معنى منطقي اقتصادي، فمثلاً إذا تحقق فرض الدخل الدائم لـ Milton Friedman's فإن قيمة الجزء الثابت المقطوع من المحور الصادي في الانحدار الخاص بالاستهلاك الدائم على الدخل الدائم يجب أن يساوى الصفر.

3 - **المتغيرات المنحدرة تكون غير مترابطة**: أي أن المتغيرات المفسرة أو المنحدرة يجب ألا تكون مرتبطة مع حد الخطأ.

(4) D. F. Hendry and J. F. Richard, "The Econometric Analysis of Economic Time Series," International Statistical Review, bol. 51, 1983, pp. 3-33.

4 - **اتساق المعالم**: بمعنى أن قيم المعالم يجب أن تكون مستقرة. وإلا سيكون التنبؤ صعبًا. فكما قال Friedman: "اختبار صلاحية النموذج الافتراضي الأفضل هو مقارن تنبؤات النموذج بالواقع".[5] في غياب اتساق المعالم لا يكون مثل هذا التنبؤ موجودًا.

5 - **ترابط البيانات**: أي أن البواقي المقدرة من النموذج، يجب أن تكون عشوائية تمامًا. بمعنى آخر إذا كان النموذج المختار مناسبًا فإن بواقي هذا النموذج يجب أن تكون عشوائية تمامًا white noise وإذا لم تكن كذلك، فإن هناك خطأ توصيف في النموذج. وسنناقش لاحقًا طبيعة أخطاء التوصيف.

6 - **النموذج الشامل**: فيكون النموذج شاملاً عندما تكون قادرًا من خلاله على توضيح نتائج باقي النماذج المحتملة. أي باختصار النماذج الأخرى لا تقدم أي جديد بعد اختيارك للنموذج الشامل.

أحيانًا نختار معيارًا ما للنموذج "الجيد" ونهمل معيارًا آخر حتى نستطيع تكوين النموذج فعليًا، وبالتالي غالبًا ما يوجد خطأ توصيف في النموذج المختار عمليًا، وسنناقش ذلك بالتفصيل في الفقرة القادمة.

2.13 أنواع أخطاء التوصيف:
TYPES OF SPECIFICATION ERRORS

افترض أنه وفقًا للمعايير السابق ذكرها، اخترنا النموذج التالي كنموذج جيد. وافترض أن هذا النموذج يمكن كتابته كالتالي:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_{1i} \qquad \textbf{(1.2.13)}$$

حيث Y = إجمالي تكلفة الإنتاج، و X = الناتج. المعادلة (1.2.13) تعتبر مثالاً شهيرًا في كتب الاقتصاد، وتسمى دالة التكلفة الكلية التكعيبية.

ولكن افترض أنه لأسباب ما (مثلاً شكل الانتشار) اختار باحث آخر النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_{2i} \qquad \textbf{(2.2.13)}$$

(5) Milton friedman, "The Methodology of Positive Economics," in Essays in Positive Economics, University of Chicago press, Chicago, 1953, p. 7.

لاحظ أننا غيرنا الرموز حتى نستطيع التفرقة بين هذا النموذج والنموذج الصحيح.

بما أن (1.2.13) يفترض صحته، فإن (2.2.13) سيكون قيمة خطأ توصيف. والخطأ الموجود هنا هو حذف متغير مهم في النموذج وهو (X_i^3) وبالتالي فإن حد الخطأ u_{2i} في (2.2.13) سيكون في الحقيقة كالتالي:

$$u_{2i} = u_{1i} + \beta_4 X_i^3 \qquad (3.2.13)$$

سنوضح لاحقًا أهمية هذه العلاقة.

والآن افترض أن باحثًا آخر اختار النموذج التالي:

$$Y_i = \lambda_1 + \lambda_2 X_i + \lambda_3 X_i^2 + \lambda_4 X_i^3 + \lambda_5 X_i^4 + u_{3i} \qquad (4.2.13)$$

إذا كان (1.2.13) صحيحًا فإن (4.2.13) يوجد فيه أيضًا خطأ توصيف. هذا الخطأ هو إضافة متغير غير ضروري وغير مهم للنموذج، وذلك حيث إنه في النموذج الصحيح يفترض أن تكون 15 تساوي الصفر. وبناءً على ذلك يكون حد الخطأ في الحقيقة كالتالي:

$$u_{3i} = u_{1i} - \lambda_5 X_i^4 \qquad (5.2.13)$$

$= u_{1i}$ since 15 = 0 in the true model (لماذا؟)

والآن افترض أن باحثًا آخر اختار النموذج التالي:

$$\ln Y_i = \gamma_1 + \gamma_2 X_i + \gamma_3 X_i^2 + \gamma_4 X_i^3 + u_{4i} \qquad (6.2.13)$$

وبالتالي بالنسبة للنموذج الصحيح، فإن هناك تحيزًا في توصيف النموذج السابق، هذا التحيز يرجع إلى استخدام شكل دالة غير صحيح: ففي (1.2.13) تكون Y خطية بينما في (6.2.13) تظهر في صورة لوغاريتمية خطية.

وأخيرًا افترض أن باحثًا آخر اختار النموذج التالي:

$$Y_i^* = \beta_1^* + \beta_2^* X_i^* + \beta_3^* X_i^{*2} + \beta_4^* X_i^{*3} + u_i^* \qquad (7.2.13)$$

حيث $Y_i^* = Y_i + \varepsilon_i$ و $X_i^* = X_i + w_i$ وبالتالي فإن ε_i و w_i فإن الباحث استخدم Y_i و X_i الصحيحة فإن الباحث استخدم Y_i^* و X_i^* والتي تشتمل على حدود أخطاء.

وبالتالي في (7.2.13) يوجد خطأ توصيف. في الواقع العملي تكون البيانات بها أخطاء راجعة للتقريب أو أخطاء راجعة إلى عدم التغطية الكاملة للدراسة أو ببساطة أخطاء لحذف بعض المفردات. العلوم الاجتماعية تعتمد عادة على البيانات الثانوية، ولا تكون لدينا وسيلة لمعرفة أنواع الأخطاء المحتملة، والتي تقوم بها المؤسسة التي جمعت البيانات الأولية.

أحد أنواع أخطاء التوصيف يعتمد على طريقة إدخال الخطأ العشوائي u_i (أو u_t) إلى النموذج. فعلى سبيل المثال، اعتبر نموذج الانحدار ثنائي المتغيرات غير المشتمل على جزء ثابت كالتالي:

$$Y_i = \beta X_i u_i \qquad \textbf{(8.2.13)}$$

حيث إن الخطأ العشوائي موجود في صورة مضروب، ويكون $\ln u_i$ مستوفيًا فروض الـ CLRM، وافترض أن لدينا أيضًا النموذج التالي:

$$Y_i = \alpha X_i + u_i \qquad \textbf{(9.2.13)}$$

وهنا يضاف حد الخطأ، على الرغم من أن المتغيرات لم تختلف في النموذجين، إلا أننا نلاحظ أن معامل الميل في (8.2.13) هو النموذج "الصحيح" أو "الحقيقي" فهل α المقدرة تعتبر مقدراً غير متحيز لـ β الحقيقية؟ أي هل ستكون $E(\hat{\alpha}) = \beta$؟ إذا لم يتحقق ذلك، يكون هناك خطأ في توصيف حد الخطأ العشوائي، والذي يعتبر أحد مصادر أخطاء التوصيف، وبالتالي لتجميع كل الملاحظات السابقة والخاصة بتكوين نموذج تطبيقي لتوصيف البيانات، **فمن الممكن أن يقع الباحث في أحد أخطاء التوصيف التالية**:

1 - حذف واحد أو أكثر من المتغيرات المهمة.

2 - إضافة واحد أو أكثر من المتغيرات غير المهمة.

3 - استخدام شكل دالة خاطئ.

4 - أخطاء القياس.

5 - توصيف غير سليم لحد الخطأ العشوائي.

قبل البدء في اختيار هذه الأنواع من أخطاء التوصيف بالتفصيل، يجب التفرقة بين أخطاء توصيف النموذج وأخطاء التوصيف غير الصحيح للنموذج. فالأنواع

الأربعة الأولى السابقة أساسية في تحديد طبيعة أخطاء توصيف النموذج، حيث إن لدينا بشكل ما تحديدًا للنموذج "الصحيح" ولكن لأسباب ما لا نقدره. في أخطاء التوصيف غير الصحيح للنموذج، لا نعرف طبيعة النموذج الصحيح حتى نأخذه كخطوة مبدئية. وهنا يمكن أن يسترجع القارئ التعارض بين الكنزيين والنقديين. فالنقديون يعطوا أولوية أساسية للمال في تفسير التغير في GDP في حين أن الكينزيين يهتمون أكثر بدور الحكومة في تفسير التغير في GDP، وبالتالي هناك نموذجان متنافسان.

في المثال التالي، سنعتبر مبدئيًا أخطاء توصيف النموذج، ثم اختيار أخطاء التوصيف غير السليم للنموذج.

3.13 عواقب أخطاء توصيف النموذج :
CONSEQUENCES OF MODEL SPECIFICATION ERRORS

بغض النظر عن مصدر أخطاء التوصيف، ما هي العواقب؟ لتبسيط المناقشة، ستتم الإجابة على السؤال السابق في إطار نموذج الانحدار ثلاثي المتغيرات ، وسنهتم بأول مصدرين للأخطاء التوصيفة السابق مناقشتها وهي بالتحديد (1) توصيف النموذج بعدد من المتغيرات أقل من العدد الصحيح، أي أن هناك حذفًا لمتغير أو أكثر من المتغيرات المهمة بالنموذج و (2) توصيف النموذج بعدد من المتغيرات أكثر من العدد الصحيح، أي أن هناك إضافة لمتغير أو أكثر غير مهم بالنموذج. مناقشتنا الحالية من الممكن بسهولة أن يتوسع نطاقها لتشمل أكثر من ثلاثة متغيرات بنموذج الانحدار، ولكنها تحتاج إلى مهارات جبرية[6]، واستخدام جبر المصفوفات يصبح ضرورة ملحة إذا وجد أكثر من ثلاثة متغيرات في نموذج الانحدار.

توصيف النموذج بأقل من الصحيح (حذف متغير مهم) :
Underfitting a Model (Omitting a Relevant Variable)

افترض أن النموذج الصحيح كالتالي :

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (1.3.13)$$

(6) ولكن انظر في تمرين (32.13) .

ولكن لأسباب ما استخدمنا النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i \quad (2.3.13)$$

عواقب حذف المتغير X_3 تكون كالتالي:

1 - إذا كان المتغير المحذوف مرتبطًا مع المتغير الموجود بالنموذج X_2، وكان معامل الارتباط r_{23} بين المتغيرين لا يساوي الصفر، فإن $\hat{\alpha}_1$ و $\hat{\alpha}_2$ ستكون مقدرات متحيزة وغير متسقة أيضًا.

وبالتالي يكون $E(\hat{\alpha}_1) \neq \beta_1$ و $E(\hat{\alpha}_2) \neq \beta_2$ وهذا التحيز لا يختفي بزيادة حجم العينة.

2 - حتى إذا لم يكن X_2 و X_3 مرتبطين فإن $\hat{\alpha}_1$ سيكون متحيزًا على الرغم من أن $\hat{\alpha}_2$ الآن سيكون غير متحيز.

3 - مقياس التباين التقليدي لـ $\hat{\alpha}_2 (= \sigma^2 / \sum x_{2i}^2)$ سيكون مقدرًا متحيزًا لتباين المعلمة الحقيقية $\hat{\beta}_2$.

4 - من توابع ذلك أيضًا أن تكون فترات الثقة اختبارات الفروض غير سليمة، وتؤدي إلى استنتاجات خاصة بمعنوية المعالم المقدرة خاطئة.

5 - أحد العواقب المهمة أيضًا أن التنبؤ بناء على النموذج غير الصحيح ستكون مضللة وغير سليمة، وأيضًا فترات الثقة المتنبأ بها.

بالطبع إثبات كل نقطة من النقاط السابقة يأخذنا بعيدًا عن إطار هذا الكتاب.[7] إلا أننا استعرضنا في الملحق A13، الفقرة 1.A13 التالي:

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{32} \quad (3.3.13)$$

حيث b_{32} هو ميل الانحدار الخاص بالمتغير المحذوف X_3 على المتغير الموجود في النموذج X_2 ($b_{23} = \sum x_{3i} x_{2i} / \sum x_{2i}^2$). كما يتضح من (3.3.13) فإن $\hat{\alpha}_2$ مقدر متحيز إلا إذا كان β_3 أو b_{32} أو كلاهما يساوي الصفر.

---

(7) للتفاصيل الجبرية انظر Jan Kmenta, Elements of Econometrics, Macmillan, New York, 1971, pp. 391-399. Those with a matrix algebra background may want to consult J. Johnston, Econometrics Methods, 4th ed., McGraw-Hill, New York, 1997, pp. 119-112.

وقد أشرنا من قبل إلى أنه إذا كانت β_3 تساوي الصفر، فإنه لا يوجد في الأساس خطأ توصيف. المعامل $b_{3\,2}$ سيساوي الصفر إذا كان X_2 و X_3 غير مرتبطتين. وهذا نادر الحدوث في أغلب البيانات الاقتصادية.

عمومًا مقدار التحيز سيعتمد على حد التحيز $\beta_3 b_{3\,2}$. فإذا كان، على سبيل المثال، β_3 موجبة (أي أن X_3 له تأثير طردي على Y) و $b_{3\,2}$ موجب (أي أن X_2 و X_3 مرتبطتين طرديًا) فإن $\hat{\alpha}_2$، في المتوسط ستقدر β_2 بأعلى من قيمتها الحقيقية (أي تحيز موجب). وهذه النتيجة متوقعة، حيث إن X_2 لا تمثل فقط التأثير المباشر على Y، ولكن أيضًا التأثير غير المباشر (عن طريق X_3) على Y. باختصار، X_2 يظهر لها دورًا مؤثرًا في النموذج من خلال تأثير X_3، والذي لا يظهر له أثر مباشر لأنه غير "مسموح" له بالظهور في النموذج. كمثال تطبيقي دعنا نسترجع المثال السابق مناقشته في الفصل (7).

مثال توضيحي : ILLUSTRATIVE EXAMPLE

وفيات الأطفال (مرة أخرى) : Child Mortality Revisited

يعمل انحدار وفيات الأطفال (CM) على GNP للفرد (PGNP) ومعدل الأمية لدى السيدات (FLR) حصلنا على نتائج الانحدار الموضحة في المعادلة (2.6.7). وفي هذه النتائج، نجد معاملات الميل الجزيئية للمتغيرين هي 0.0056- و 2.2316- بالترتيب. ولكن إذا حذفنا متغير FLR، نحصل على النتائج المعطاة في المعادلة (2.7.7) إذا اعتبرنا أن (2.6.7) هو النموذج الصحيح، إذن (2.7.7) يعاني من خطأ التوصيف، حيث إنه تم حذف متغير مهم وهو FLR. ترى الآن أنه في النموذج الصحيح معامل PGNP كان 0.0056- في حين أنه في النموذج "غير السليم" (2.7.7) كان 0.0114-.

باستخدام القيم المطلقة، نرى أن PGNP الآن له تأثير أكبر على CM مقارنة مع النموذج الصحيح. ولكن إذا قمنا بعمل انحدار لـ FLR على PGNP (انحدار للمتغير المحذوف على المتغير الباقي في النموذج). معامل الميل في هذا الانحدار [$b_{3\,2}$ في المعادلة (3.3.13)] يساوي 0.00256.[8] وهذا يعني أنه كلما زاد PGNP بوحدة واحدة فإنه في المتوسط تزداد FLR بحوالي 0.00256 وحدة. ولكن إذا زادت FLR بثلاث وحدات فإن تأثير ذلك على CM يكون كالتالي:

$$(-2.2316)\,(0.00256) = \hat{\beta}_3 b_{3\,2} = -\,0.00543$$

(8) نتائج الانحدار كالتالي :

$$\widehat{FLR} = 47.5971 \; + \; 0.00256PGNP$$
$$se = (3.5553) \quad (0.0011) \qquad r^2 = 0.0721$$

وبالتالي من (3.3.13) لدينا في النهاية

$$(\hat{\beta}_2 + \hat{\beta}_3 b_{23}) = -0.0056 + (-2.2316)(0.00256) \approx -0.0111$$

وهذه نفس القيمة تقريبًا الخاصة بمعامل PGNP التي نحصل عليها من النموذج غير الصحيح (2.7.7) .[9]

كما يتضح من هذا المثال، التأثير الحقيقي لـ PGNP على CM أقل بـ من (0.0056 –) من الأثر المقترح من النموذج غير الصحيح (2.7.7) والمساوي (0.0114 –) .

دعنا الآن نختبر تباينات $\hat{\alpha}_2$ و $\hat{\beta}_2$.

$$\operatorname{var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \tag{4.3.13}$$

$$\operatorname{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1 - r_{23}^2)} = \frac{\sigma^2}{\sum x_{2i}^2}\text{VIF} \tag{5.3.13}$$

حيث VIF (مقياس للتعدد الخطي) هوتباين معامل التضخم $[=1/(1 - r_{23}^2)]$ والذي سبق وناقشناه في الفصل (10) و r_{23} هو معامل الارتباط بين X_2 و X_3. المعادلات (4.3.13) و (5.3.13) سبق وتعرضنا لها في الفصلين 3 و7.

بما أن المعادلات (4.3.13) و (5.3.13) غير متطابقة، فإنه في العموم تباين $(\hat{\alpha}_2)$ سيكون مختلفًا عن تباين $(\hat{\beta}_2)$. ولكننا نعلم أن $\operatorname{var}(\hat{\beta}_2)$ غير متحيز (لماذا؟). وبالتالي فإن $\operatorname{var}(\hat{\alpha}_2)$ متحيز، وهذا يدعم ما سبق وذكرناه في النقطة (4) سابقًا.

بما أن $0 < r_{23}^2 < 1$ فإنه $\operatorname{var}(\hat{\alpha}_2) < \operatorname{var}(\hat{\beta}_2)$. وهنا لدينا مشكلة وهي كالتالي: على الرغم من أن $\hat{\alpha}_2$ مقدر متحيز، إلا أن تباينه أقل من تباين المقدر غير المتحيز (بالطبع تم استبعاد الحالة التي يكون فيها $r_{23} = 0$، بما أنه في الواقع يكون هناك بعض الارتباط بين المتغيرات المنحدرة) وبالتالي لابد من المفاضلة بين الأمرين السابقين.[10]

الموضوع لم ينته عند هذا الحد، فالقيمة المقدرة لـ σ^2 من النموذج (2.3.13) والأخرى المقدرة من النموذج الصحيح (1.3.13) غير متساويين حيث إن RSS للنموذجين وكذلك درجات الحرية الخاصة بهما مختلفة. تذكر أننا حصلنا على

(9) لاحظ أنه في النموذج الصحيح $\hat{\beta}_2$ و $\hat{\beta}_3$ مقدرات غير متحيزة لقيمهم الحقيقية .

(10) للتغلب على مسألة المفاضلة بين التحيز والكفاءة، يمكن للباحث أن يختار تصغير مربعات متوسطات الخطأ (MSE)، حيث إنه يتعامل مع التحيز والكفاءة في نفس الوقت. لمزيد من التفاصيل عن MSE، انظر الملحق الإحصائي **.App.A** . انظر أيضًا تمرين 6.13.

تقدير لـσ^2 كالتالي $\hat{\sigma}^2 \neq RSS/df$ وبالتالي فالتقدير يعتمد على عدد المتغيرات المنحدرة الموجودة في النموذج، بالإضافة إلى درجات الحرية (= n، عدد المعالم المقدرة). والآن إذا أضفنا متغيرات جديدة للنموذج، فإن RSS بوجه عام يقل (تذكر أنه كلما زاد عدد المتغيرات المضافة إلى النموذج، فإن R^2 يزداد) ولكن درجات الحرية أيضًا تقل، حيث عدد المعالم المقدرة يزيد. المحصلة النهائية تعتمد على ما إذا كان RSS يقل بشكل كفء أم لا مما يعوض مشكلة تقليل درجات الحرية الناتجة عن إضافة متغيرات منحدرة جديدة.

من المحتمل أنه مثلاً إذا كان لمتغير منحدر تأثير كبير على المتغير المنحدر عليه، فإنه يؤدي إلى تقليل RSS أقل من الخسارة في درجات الحرية كنتيجة لإضافة المتغيرات إلى النموذج. إضافة مثل هذا المتغير إلى النموذج لن يؤدي إلى تقليل التحيز فقط، وإنما سيزيد أيضًا من الدقة في التقدير (أي سيقلل الأخطاء المعيارية).

على الجانب الآخر، إذا كان المتغير له تأثير محدود على المتغير المنحدر، وإذا كانوا مرتبطين بدرجة عالية (أي أن VIF كبير) يمكننا تقليل التحيز في معاملات المتغيرات الموجودة فعليًا في النموذج، ولكن يزداد الخطأ المعياري لهم (أي سيكون أقل كفاءة).

بالطبع المفاضلة في هذه الحالة بين التحيز والدقة تكون مهمة وحيوية. وكما ترى من المناقشة السابقة، هذه المفاضلة ستعتمد على الأهمية النسبية للمتغيرات المنحدرة المختلفة.

يمكن أن نستنتج التالي من المناقشة السابقة: دعنا نعتبر الحالة الخاصة التي يكون فيها $r_{23} = 0$ أي أن X_2 و X_3 غير مرتبطين. سيؤدي ذلك إلى أن يكون $b_{3\,2}$ يساوي الصفر (لماذا؟). وبالتالي نرى من (3.3.13) أن $\hat{\alpha}_2$ الآن مقدر غير متحيز.[11] ويبدو أيضًا من (4.3.13) و (5.3.13) أن تباينات $\hat{\alpha}_2$ و $\hat{\beta}_2$ متساوية.

هل لا يوجد أي ضرر من حذف المتغير X_3 من النموذج حتى وإذا كان مهمًا من الناحية النظرية؟ الإجابة بوجه عام هي لا. ولكن في هذه الحالة كما سبق وذكرنا، فإن $\text{var}(\hat{\alpha}_2)$ المقدر من (4.3.13) مازال متحيزًا وبالتالي اختبارات الفروض ستكون

(11) لاحظ أنه على الرغم من أن $\hat{\alpha}_1$ مازال متحيزًا فإننا نفرض أن $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}_2 - \hat{\beta}_3\bar{X}_3$ ، حيث $\hat{\alpha}_1 = \bar{Y} - \hat{\alpha}_2\bar{X}_2$ حتى إذا كان $\hat{\alpha}_2 = \hat{\beta}_2$. فإن مقدرا الجزءين الثابتين لن يتساويان.

نتائجها غير موثوق فيها.[12] بالإضافة إلى أنه في العديد من الأبحاث الاقتصادية يكون X_2 و X_3 مرتبطين مما يؤدي إلى المشكلة السابق ذكرها.

وبالتالي، فالنقطة المهمة هنا هي: إذا تم تكوين النموذج على أساس نظري سليم، يكون حذف أي متغير من النموذج أمرًا لا ينصح به.

إدخال متغير غير ضروري في النموذج (توفيق النموذج بعدد من المتغيرات أكثر من المطلوب):

Inclusion of an Irrelevant Variable (Overfitting a Model)

دعنا الآن نفترض التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \qquad \textbf{(6.3.13)}$$

هو النموذج الصحيح، ولكننا استخدمنا النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_i \qquad \textbf{(7.3.13)}$$

وهذا يمثل خطأ توصيف ناتج عن إضافة متغير غير مهم إلى النموذج.

عواقب خطأ التوصيف السابق كالتالي:

1 - مقدرات OLS لمعالم النموذج "غير الصحيح" تكون جميعها غير متحيزة ومتسقة. أي أن $E(\alpha_1) = \beta_1$ ، $E(\hat{\alpha}_2) = \beta_2$ و $E(\hat{\alpha}_3) = \beta_3 = 0$.

2 - تباين الخطأ σ^2 مقدر بشكل سليم.

3 - فترات الثقة العادية واختبارات الفروض سليمة.

4 - عمومًا مقدرات α's ستكون غير كفء أي أن تباينها عمومًا أكبر من نظيرها الخاص بـ $\hat{\beta}$'s للنموذج الصحيح.

إثبات بعض هذه النقاط موجود في ملحق A13 الفقرة 2.A13 . النقطة المهمة هنا هي الكفاءة النسبية لـ $\hat{\alpha}$'s والتي يمكن توضيحها بسهولة كالتالي:

من الـ OLS التقليدية نعلم أن:

$$\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \qquad \textbf{(8.3.13)}$$

(12) لمزيد من التفاصيل، انظر Adrian C. Darnell, A Dictionary of Econometrics, Edward Elgar Publisher, 1994, pp. 371-372.

و

$$\text{var}(\hat{\alpha}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} \qquad (9.3.13)$$

وبالتالي :

$$\frac{\text{var}(\hat{\alpha}_2)}{\text{var}(\hat{\beta}_2)} = \frac{1}{1 - r_{23}^2}$$

بما أن $0 \leq r_{23}^2 \leq 1$ ، يكون $\text{var}(\hat{\alpha}_2) \geq \text{var}(\hat{\beta}_2)$ أي أن تباين $\hat{\alpha}_2$ أكبر من تباين $\hat{\beta}_2$ وإن كان في المتوسط $\hat{\alpha}_2 = \hat{\beta}_2$ (أي أن $E(\hat{\alpha}_2) = \beta_2$) معنى هذه النتيجة أن إدخال متغير غير مهم مثل X_3 إلى النموذج جعل تباين $\hat{\alpha}_2$ أكبر من المفروض ، وبالتالي قلت دقة $\hat{\alpha}_2$. وهذا ينطبق أيضًا على $\hat{\alpha}_1$.

لاحظ عدم التماثل بين نوعي الخطأ السابق ذكرهما من أخطاء التوصيف . فإذا حذفنا متغيرًا مهمًا وضروريًا ، فإن معاملات المتغيرات الباقية في النموذج ستكون بوجه عام متحيزة وغير متسقة ، وتباين الخطأ سيقدر بشكل غير سليم ، واختبارات الفروض تكون غير سليمة . على الجانب الآخر ، إضافة متغير غير مهم أو غير ضروري إلى النموذج ستظل التقديرات غير متحيزة ومتسقة ، وتباين الخطأ سيقدر بشكل سليم ، وكل خطوات ونتائج اختبارات الفرض التقليدية تكون سليمة . المشكلة الوحيدة لإدخال نموذج غير مهم إلى النموذج أن التباين المقدر للمعاملات سيكون أكبر من المفروض ، وبالتالي كنتيجة لذلك استبدلنا الإحصائي عن المعالم سيكون له دقة أقل .

وليس من المطلوب الآن أن تستنتج عزيزى القارئ أنه من الأفضل إضافة متغير غير مهم إلى النموذج عن حذف متغير مهم ، حيث إن إضافة هذا المتغير غير المهم ستؤدي إلى خسارة في كفاءة المقدرات ، وقد تؤدي إلى مشكلة التعدد الخطي (لماذا؟) بالإضافة إلى مشكلة تقليل درجات الحرية وبالتالي :

عمومًا ، الأسلوب الأفضل هو إضافة المتغيرات المفسرة إلى النموذج على أساس نظري ، وتكون هذه المتغيرات مؤشرة بشكل مباشر على المتغير التابع وتأثيرها لا يكفي تضمينه في المتغيرات الأخرى الموجودة فعلاً في النموذج .[13]

---

(13) Michael D. Intriligator, Econometric Models, Techniques and Applications, Prentice Hall, Englewood Cliffs, N.J., 1978, p. 189. Recall the Occam's razor principle.

4.13 اختبارات لأخطاء التوصيف :
TESTS OF SPECIFICATION ERRORS

معرفة عواقب أخطاء التوصيف تعتبر موضوعًا قائمًا بذاته ومختلفًا عن اكتشاف خطأ التوصيف. في الأساس غالبًا ما يظهر تحيز التوصيف بدون قصد أو بدون عمد. ويكون بسبب رغبتنا في تكوين نموذج دقيق بأكبر قدر ممكن، معتمدين على أن النظرية ضعيفة أو ليس لدينا البيانات الكافية لاختبار مثل هذا النموذج. وكما لاحظ Davidson "نظرًا للطبيعة غير القابلة للتجربة للاقتصاد، لا تستطيع الجزم بكيفية تجميع البيانات. فاختبر أي فرض في الاقتصاد يعتمد في الأساس على المزيد من الفروض التي تجعلنا قادرين على تكوين نموذج مناسب، مع العلم بأن هذه الفروض قد تكون منطقية أو لا".[14]

السؤال العملي الآن ليس عن تحليل وجود أخطاء التوصيف، ولكن الأهم هو كيف يمكن اكتشافها. فمجرد التأكد من حدوث خطأ في التوصيف، فإن طرق العلاج يمكن اتباعها لتجنب العواقب. فمثلاً العلاج الواضح لمشكلة خطأ التوصيف الناتجة عن حذف متغير مهم من النموذج هو إضافة هذا المتغير إلى النموذج مباشرةً بعد التأكد من وجود بيانات خاصة به.

في الفقرة التالية، سنناقش بعض الاختبارات التي يمكن استخدامها لاكتشاف وجود خطأ التوصيف.

اكتشاف وجود متغير غير مهم في النموذج (توفيق النموذج بعدد من المتغيرات أكثر من المطلوب)
Detecting The Presence of Unnecessary Variables (Overfitting a Model)

افترض أن لدينا نموذجًا له k متغير لتفسير ظاهرة ما كالتالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i \qquad (1.4.13)$$

عمومًا، نحن غير متأكدين من أن، مثلاً، المتغير X_k فعلاً يجب أن يتواجد في هذا النموذج. أحد الطرق البسيطة لاكتشاف ذلك هو اختبار معنوية القيمة المقدرة لـ β_k باختبار t التقليدي $t = \hat{\beta}_k / \mathrm{se}(\hat{\beta}_k)$. ولكن افترض أننا غير متأكدين من أهمية وجود X_3 و X_4 معًا في النموذج. يمكن اختبار ذلك مباشرة باختبار F السابق مناقشته في الفصل (8). وبالتالي اكتشاف وجود متغير غير مهم في النموذج ليس عملية صعبة.

(14) James Davidson, Econometric Theory, Blackwell Publishers, Oxford, U.K., 2000, p. 153.

ولكن من المهم تذكر أنه حتى يمكنك القيام بمثل هذه الاختبارات، يجب أن يكون هناك في الأساس نموذج للتعامل معه. فنحن نقبل هذا النموذج أولاً ثم نتعامل معه. وبالتالي، فإنه وفقًا لهذا النموذج يمكن تحديد ما إذا كان واحد أو أكثر من المتغيرات المنحدرة غير مهم في النموذج باستخدام اختبارات t و F التقليدية.

ولكن لاحظ أننا لا نستخدم اختبارات t و F لبناء النموذج، فلا يجب أن نستنتج أن Y مرتبطة بـ X_2 مثلاً لأن $\hat{\beta}_2$ لها معنوية إحصائية ثم تبدأ في توسيع النطاق وإدخال X_3 للنموذج وإبقائه فيه إذا كانت $\hat{\beta}_3$ لها معنوية إحصائية وهكذا.

فهذه الطريقة في بناء النموذج تسمى أسلوب الأعلى - الأسفل (حيث نبدأ بنموذج أصغر ويتسع نطاقه تدريجيًا) أو تسمى أيضًا التنقيب عن البيانات (مسميات أخرى مثل انحدار الأسماك، تجميع البيانات).

الهدف الأساس للتنقيب عن البيانات هو تكوين النموذج "الأفضل" بعد عمل العديد من اختبار التشخيص، وبالتالي يكون النموذج النهائي المختار هو نموذج "جيد" بمعنى أن كل المعاملات المقدرة لها الإشارة "الصحيحة" ولهم معنوية إحصائية بناء على اختبارات t و F، وقيمة R^2 تكون كبيرة بدرجة معقولة و Durbin-Watsond لها قيمة مقبولة (قريبة من 2) وهكذا. أما الجانب الفني الخاص بتطبيق تنقيب البيانات فنجده في مقولة william pool التالية "... الاعتماد على التطبيق العملي بدلاً من التطبيق نظرية اقتصادية سيكون دائمًا خطر".[15] أحد أسباب ادانة تنقيب البيانات كالتالي:

مستوى المعنوية الحقيقي والاسمى في حالة استخدام تنقيب البيانات

Nominal versus True Level of Significance in the Presence of Data Mining

خطر استخدام تنقيب البيانات والذي قد يواجه الباحث هو مستوى المعنوية التقليدي (α) مثل 1، 5 أو 10% فلن يكون هو مستوى المعنوية الحقيقي. وقد اقترح Lovell أنه إذا كان هناك من المتغيرات المنحدرة وتم اختيار k كعدد نهائي ($k \leq c$) بناء على أسلوب تنقيب البيانات، فإن مستوى المعنوية الحقيقي (α^*) مرتبط بمستوى المعنوية الاسمي (α) كالتالي:[16]

(15) William Pool, "Is Inflation Too Low", the Cato Journal, vol. 18, no. 3 Winter 1999, p. 456.
(16) M. Lovell, "Data Mining", Review of Economics and Statistics, vol. 65, 1985, pp. 1-12.

$$\alpha^* = 1 - (1-\alpha)^{c/k} \quad (2.4.13)$$

أو تقريبًا كالتالي :

$$\alpha^* \approx (c/k)\alpha \quad (3.4.13)$$

فمثلاً إذا كان $c = 15$ ، $k = 5$ و $\alpha = 5\%$ فمن (3.4.13) يكون مستوى المعنوية الحقيقي هو 15% = (5) (15/5) . وبالتالي إذا استخدم الباحث تنقيب البيانات واختار 5 من 15 متغيرًا منحدرًا واستعرض نتائجه على أساس أنه استخدم مستوى معنوية 5% ، واستعرض بعد ذلك نتائج المعنوية الإحصائية ، يجب علينا أن ندرك أن هذه النتائج غير سليمة بالمرة ، حيث إننا نعلم أن مستوى المعنوية (الحقيقي) هو 15% . ويجب ملاحظة أنه إذا كان $c = k$ فلا يوجد تنقيب للبيانات ، ومستوى المعنوية الحقيقي والاسمي متساويان . بالطبع في الواقع معظم الباحثين يسجلون نتائجهم الخاصة بالانحدار "النهائي" . بدون توضيح استخدم أسلوبًا لتنقيب البيانات أو الخطوات المختلفة التي قاموا بها للوصول إلى هذا النموذج .[17]

وعلى الرغم من العيوب الواضحة لهذه الطريقة ، فإنه هناك زيادة عملية تم استخدامها خصوصًا في الاقتصاد القياسي التطبيقي . وبالتالي يصبح أسلوب الوصول إلى نموذج (بدون استخدام تنقيب البيانات) من الصعب الحفاظ عليه . وكما لاحظ Zaman :

للأسف التجارب المختلفة بيانات حقيقية أوضحت أن [الأسلوب الذي لا يعتمد على تنقيب البيانات] غير متاح وغير مرغوب فيه أيضًا . غير متاح لأنه من النادر في نظرية الاقتصاد أن تجد ما يشير إلى نموذج فريد لتوصيف البيانات . وغير مرغوب فيه ، لأن بعض المفاهيم التي يتعلم الباحث من البيانات لاستخدام أو عدم استخدام نماذج معينة لا يمكن الاعتماد عليها في هذا الأسلوب ، حتى ولو على سبيل الصدفة . كان النموذج المبدئي نموذجًا جيدًا . يكون من المفيد البحث عن أنواع من النماذج الأخرى التي قد تتفق أو تختلف مع البيانات محل الدراسة .[18]

(17) لمزيد من التفاصيل وأساسيات الموضوع انظر: Wallace, T. D., "Pretest Estimation in Regression: A Survey", American Journal of Agricultural economics, vol. 59, 1977, pp. 431-443.

(18) Asad Zaman, Statistical Foundations for Econometric Techniques, Academic Press, New York, 1996, p. 226.

وقد اتفق Kerry Patlerson مع ذلك ووضح التالي :

أسلوب [تنقيب البيانات] يقترح أن يتم الفصل بين النظرية الاقتصادية من جهة ، وبين التطبيق العلمي للبيانات من جهة أخرى .[19]

وبدلاً من الاستمرار في المقارنة بين الأسلوب التقليدي (الذي لا يوجد فيه تنقيب للبيانات] وأسلوب التنقيب في البيانات ، يمكن للباحث أن يستخدم وجهة نظر Peter Kennedy والتي تنص على التالي :

[توصيف النموذج] يجب أن يتم وفقًا للمزج بين النظرية والبيانات ، والاختبارات المختلفة المستخدمة للتوصيف ، يجب أن تستخدم لتقليل تكلفة تنقيب البيانات . أمثلة على تلك الأساليب هي : تجنب البيانات التي لا ينطبق عليها اختبارات التنبؤ ، تعديل مستويات المعنوية [a la lovell] وتجنب الطرق المشكوك فيها مثل تكبير R^2 .[20]

إذا اقتصرنا النظر لتنقيب البيانات على أنه أسلوب يعتمد على بيانات عملية قد يكون بها أخطاء ، وقد يكون بها فائدة كبيرة لتوصيف النموذج يكون تنقيب البيانات في ذلك الحين أسلوبًا فعالاً جداً وينصح باستخدامه . وفقًا لـ kennedy مرة أخرى ، " فن الاقتصاد القياسي التطبيقي يعتمد على اختيار البيانات وفقًا للنظرية ، ومحاولة تفادي الأخطار المحتملة للتنقيب في البيانات " .[21]

اختبارات للمتغيرات المحذوفة وشكل الدالة غير الصحيح :

Tests for Omitted Variables and Incorrect Functional Form

في الواقع لا تستطيع الجزم بأن النموذج المختار وفقًا لاختبارات تطبيقه هو " النموذج السليم ولا شيء غير ذلك " . فبناء على النظرية والجانب التطبيقي السابق في نفس المجال ، نصل إلى نموذج نعتقد أنه يتناول كافة جوانب الموضوع محل الدراسة . ثم نطبق هذا النموذج باستخدام اختبارات تطبيقية . بعد أن نحصل على النتائج ، نبدأ في تحليل البيانات واضعين في الاعتبار الأساليب التي تم بناء على

(19) Kerry Patterson, An Introduction to Applied Econometrics, St. Martin's Press, New York, 2000, p. 10.

(20) Peter Kennedy, "Sinning in the Basement: What Are the Rules? The Ten Commandments of Applied Econometrics," unpublished manuscript.

(21) Kennedy, op. cit., p. 13.

اختيار النموذج الجيد والسابق ذكرها مباشرةً. عند هذه المرحلة، نحن نحاول معرفة درجة دقة النموذج المختار. لتحديد دقة النموذج، تستعرض بعض النتائج مثل قيمة R^2، قيم t المقدرة إشارات المعاملات المقدرة وعلاقتها بتوقعاتنا السابقة، إحصاء Durbin-Watson صغير جداً، هنا يجب أن نقلق على جودة ودقة النموذج ونفكر في أساليب علاجية: فقد نكون حذفنا متغيراً مهماً، أو استخدمنا شكل دالة غير سليم، أو لم تستخدم الفروق وكانت لدينا بيانات سلسلة زمنية (لإزالة أثر الارتباط التسلسلي) وهكذا. لمعرفة ما إذا كان النموذج المختار يعاني من واحد أو أكثر من هذه المشاكل، يمكن أن تستخدم بعض الأساليب التالية.

اختبار البواقي: Examination of Residuals

كما سبق وأشرنا في الفصل (12)، اختبار البواقي يعتبر أداة تشخيص بصرية مهمة لاكتشاف الارتباط الذاتي أو اختلاف التباين. ولكن هذه البواقي يمكن اختبارها أيضًا، خصوصًا في البيانات المقطعية، وبالنسبة لأخطاء التوصيف المتعلقة بحذف أحد المتغيرات المهمة أو استخدام شكل دالي غير سليم. إذا كان هناك أحد هذه الأخطاء، فإن الرسم البياني الخاص بالبواقي سيكون له نمط محدد.

للتوضيح، دعنا نعتبر دالة التكلفة التكعيبية والتي سبق وناقشناها في الفصل (7). افترض أن دالة التكلفة الكلية الحقيقية كالتالي حيث Y = التكلفة الكلية و X = الناتج:

$$Y_i + \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \beta_4 X_i^3 + u_i \quad (4.4.13)$$

ولكن إذا قام البحث باستخدام الدالة التربيعية التالية

$$Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + u_i^2 \quad (5.4.13)$$

وإذا افترضنا أن باحثًا آخر اختار الدالة الخطية التالية:

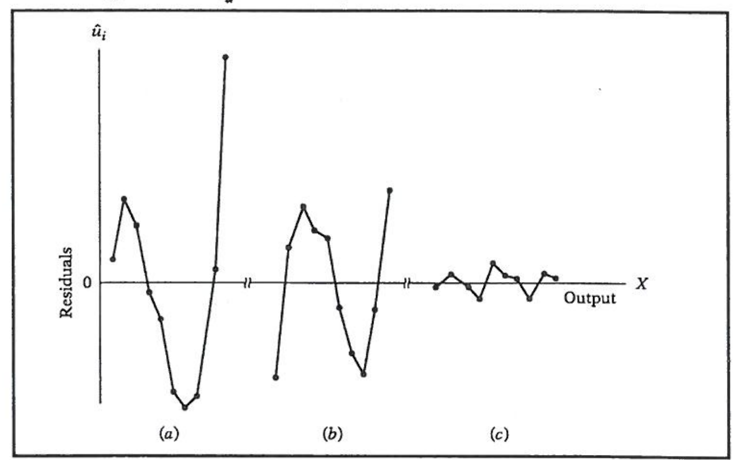
$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \quad (6.4.13)$$

على الرغم من أننا نعلم أن كلاً من الباحثين السابقين قد وقعا في خطأ التوصيف، فدعنا نستعرض شكل البواقي المقدرة في كل من النماذج الثلاثة السابقة. [بيانات التكلفة – النتائج معطاة في جدول (4.7)]. شكل (1.13) يتضح منه التالي: كلما انتقلنا من اليسار إلى اليمين، أي تقترب من الحقيقة، ليس فقط تقل البواقي (بصورة مطلقة) ولكن أيضًا يختفي الشكل أو النمط الدائري المصاحب للنماذج غير الصحيحة.

وبالتالي، فائدة استخدام شكل البواقي واضحة تمامًا. فإذا كان هناك خطأ توصيف، فإن البواقي سيظهر فيها نمط محدد واضح.

اختبار d لـ Durbin-Watson مرة أخرى: The Durbin-Watson d Statistic Once Again

إذا اخترنا إحصاء Durbin-Watson d المحسوب من جدول (1.13) نرى أنه يساوي 0.716 بالنسبة لدالة التكلفة الخطية، مما يعني أن هناك "ارتباطًا" طرديًا في البواقي المقدرة: بالنسبة لـ $n = 10$ و $k' = 1$ فإن الـ 5% قيمة حرجة لـ d هي $d_L = 0.879$ و $d_U = 1.320$.



شكل (1.13) البواقي $\hat{u}_i$ محسوبة بناءً على (a) دالة تكلفة خطية، (b) دالة تكلفة تربيعية و (c) دالة تكلفة تكعيبية

جدول (1.13) البواقي المقدرة من دالة التكلفة الكلية الخطية، التربيعية والتكعيبية

Observation number	$\hat{u}_i$, linear model*	$\hat{u}_i$, quadratic model†	$\hat{u}_i$, cubic model**
1	6.600	−23.900	−0.222
2	19.667	9.500	1.607
3	13.733	18.817	−0.915
4	−2.200	13.050	−4.426
5	−9.133	11.200	4.435
6	−26.067	−5.733	1.032
7	−32.000	−16.750	0.726
8	−28.933	−23.850	−4.119
9	4.133	−6.033	1.859
10	54.200	23.700	0.022

* $\hat{Y}_i = 166.467 + 19.933X_i$ $R^2 = 0.8409$
(19.021) (3.066) $\bar{R}^2 = 0.8210$
(8.752) (6.502) $d = 0.716$

† $\hat{Y}_i = 222.383 - 8.0250X_i + 2.542X_i^2$ $R^2 = 0.9284$
(23.488) (9.809) (0.869) $\bar{R}^2 = 0.9079$
(9.468) (−0.818) (2.925) $d = 1.038$

** $\hat{Y}_i = 141.767 + 63.478X_i - 12.962X_i^2 + 0.939X_i^3$ $R^2 = 0.9983$
(6.375) (4.778) (0.9856) (0.0592) $\bar{R}^2 = 0.9975$
(22.238) (13.285) (−13.151) (15.861) $d = 2.70$

وبالمثل فإن قيمة d المحسوبة بناء على دالة التكلفة التربيعية هي 1.038 وتكون الـ 5% قيم حرجة هي $d_L = 0.697$ و $d_U = 1.641$ مما يعني عدم القدرة على اتخاذ قرار في هذه الحالة. ولكن إذا استخدمنا اختبار d المعدل (انظر الفصل 12) سنجد أن هناك "ارتباطًا" طرديًا في البواقي. حيث إن قيمة d المحسوبة أقل من d_U. أما بالنسبة لدالة التكلفة التكعيبية، التوصيف الصحيح، فإن قيمة d المقدرة لا تشير إلى أي "ارتباط" طردي في البواقي.[22]

"الارتباط" الموجب المشاهد في البواقي، عندما تستخدم نموذجًا خطيًا أو تربيعيًا ليس دليلاً على وجود ارتباط ذاتي (من الدرجة الأولى) في البواقي ولكن دليل على وجود خطأ في توصيف (النموذج). الارتباط المشاهد ببساطة يعبر عن المتغيرات المهمة التي لم يتم استخدامها في النموذج، وتوجد في الخطأ، ويجب فصلها من حد الخطأ وتقديمها كمتغيرات مفسرة في النموذج: فمثلاً إذا حذفنا X_i^3 من دالة التكلفة كما في (3.2.13)، فإن حد الخطأ في النموذج غير الصحيح (2.2.13) هو في الواقع يساوي $(u_{1i} + \beta_4 X_i^3)$ وسيظهر فيها نمط محدد (أي الارتباط الذاتي الموجب)، وذلك دليل على أن X_i^3 تؤثر معنويًا على الـ Y.

لاستخدام اختبار Durbin-Watson لاكتشاف خطأ التوصيف نقوم بالتالي:

1 - من النموذج المفترض، نحصل على بواقي الـ OLS.

2 - إذا اعتقدنا أن النموذج يعاني من خطأ التوصيف، حيث تم حذف متغير مفسر مهم، مثلاً Z من النموذج. لاحظ أن: المتغير Z قد يكون أحد متغيرات X المتضمنة في النموذج المفترض، أو قد تكون دالة ما في هذا المتغير مثل X_2 أو X_3.

3 - احسب الإحصاء d من البواقي بناء على المعادلة التقليدية التالية:

$$d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$$

لاحظ أن: الترميز t يشير إلى المشاهدة وليس بالضرورة يعني أن البيانات تعتبر سلسلة زمنية.

4 - من جداول Durbin-Watson، إذا كانت قيمة d المقدرة معنوية، فإنه من الممكن قبول الفرض الخاص بوجود خطأ توصيف في النموذج، وإذا حدث ذلك فعلاً، بالتالي الطرق العلاجية تقترح نفسها بوضوح كالتالي:

---

(22) في السياق الحالي، قيمة $d = 2$ تعني عدم وجود خطأ في التوصيف. (لماذا؟)

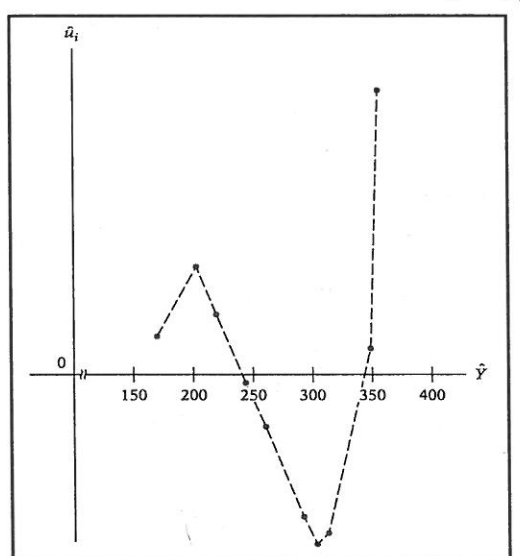
في مثالنا الحالي، المتغير Z(=X) (الناتج) كان مرتبًا بالفعل[23] وبالتالي، لا توجد ضرورة لحساب إحصاء d مجددًا. فقد رأينا أن إحصاء d لكل من دوال التكلفة الخطية والتربيعية يقترح وجود أخطاء في التوصيف. العلاج في هذه الحالة واضح ويكون كالتالي: قم بادخال القيم التربيعية والتكعيبية في دالة التكلفة الخطية، وادخل القيم التكعيبية على دالة التكلفة التربيعية. أو باختصار استخدم نموذج التكلفة التكعيبية.

اختبار RESET لـ Ramsey :

اقترح Ramsey اختبارًا عامًا لخطأ التوصيف وسماه RESET (اختبار خطأ التوصيف للانحدار Regression Specification error test)[24] سنستعرض هنا الشكل الأبسط لهذا الاختبار. لتوضيح الفكرة دعنا نستخدم مرة أخرى مثال التكلفة - الناتج، ونفترض أن دالة التكلفة هي دالة خطية في الناتج كالتالي:

$$Y_i = \lambda_1 + \lambda_2 X_i + u_{3i} \qquad (6.4.13)$$

حيث Y = التكلفة الكلية و X = الناتج. الآن إذا رسمنا البواقي $\hat{u}_i$ التي حصلنا عليها من هذا الانحدار ضد $\hat{Y}_i$ القيمة المقدرة لـ Y_i من النموذج، سنحصل على الشكل الموجود في (2.13).



شكل (2.13) البواقي $\hat{u}_i$ و Y المقدرة من دالة التكلفة الخطية: $Y_i = \lambda_1 + \lambda_2 X_i + u_i$

(23) لا يهم ما إذا قمنا بترتيب $\hat{u}_i$ بناء على X_i^2 أو X_i^3 حيث إن هذه الدوال في X مرتبة بالفعل.

(24) J. B. Ramsey, "Tests for Specification Errors in Classical Linear Least Squres Regression Analysis, "Journal of the Royal Statistical Society, series B, vol. 31, 1969, pp. 350-371.

على الرغم أن $\Sigma\hat{u}_i$ و$\Sigma\hat{u}_i\hat{Y}_i$ بالضرورة يساويان الصفر (لماذا؟ انظر الفصل 3)، إلا أن البواقي تظهر في علاقتها بـ $\hat{Y}_i$ نمط محدد في الشكل البياني الخاص بهما، مما يعني أننا إذا استطعنا استخدام $\hat{Y}_i$ في شكل ما كمتغيرات محددة في (6.4.13) فإن ذلك سيؤدي إلى زيادة R^2، وإذا كانت هذه الزيادة معنوية إحصائيًا (على أساس اختبار F السابق مناقشته في الفصل 8) فإن ذلك يعني أنه كان من الخطأ استخدام دالة التكلفة الخطية (6.4.13) وهذه هي الفكرة وراء استخدام RESET. خطوات الـ RESET تكون كالتالي:

1 – من النموذج المختار، مثل (6.4.13)، نحصل على رقم Y_i المقدرة أي $\hat{Y}_i$.

2 – نعيد عمل (6.4.13) مع تقديم $\hat{Y}_i$ للنموذج في بعض الصور الدالية المختلفة كمتغير أو متغيرات منحدرة من الشكل (2.13)، نرى أن هناك علاقة غير خطية بين $\hat{u}_i$ و $\hat{Y}_i$ مما يعني إمكانية استخدام $\hat{Y}_i^2$ و$\hat{Y}_i^3$ كمتغيرات منحدرة إضافية. وبالتالي تقوم بعمل الانحدار التالي:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 \hat{Y}_i^2 + \beta_4 \hat{Y}_i^3 + u_i \qquad \textbf{(7.4.13)}$$

3 – ارمز R^2 التي حصلت عليها من (7.4.13) بـ R^2_{new} والتي حصلت عليها من (6.4.13) ارمز لها بـ R^2_{old}. وبالتالي تستطيع استخدام اختبار F الذي سبق وقدمناه في (18.5.8) وهو كالتالي:

$$F = \frac{(R^2_{\text{new}} - R^2_{\text{old}}) / \text{عدد المتغيرات المنحدرة}}{(1 - R^2_{\text{new}}) / (n - \text{عدد المعالم في النموذج الجديد})} \qquad \textbf{(18.5.8)}$$

ونرى ما إذا كانت الزيادة في R^2 باستخدام (7.4.13) لها معنوية إحصائية.

4 – إذا كانت قيمة F المحسوبة معنوية، مثلاً عند مستوى المعنوية 5% يمكن قبول الفرض القائل بأن النموذج (7.4.13) نموذج غير صحيح.

بالعودة إلى مثالنا التوضيحي، لدينا النتائج التالية (الأخطاء المعيارية موجودة بين الأقواس).

$$\hat{Y}_i = 166.467 + 19.933X_i \qquad \textbf{(8.4.13)}$$
$$(19.021) \quad (3.066) \qquad R^2 = 0.8409$$

$$\hat{Y}_i = 2140.7223 + 476.6557X_i - 0.09187\hat{Y}_i^2 + 0.000119\hat{Y}_i^3$$
$$(132.0044) \quad (33.3951) \quad (0.00620) \quad (0.0000074) \qquad \textbf{(9.4.13)}$$
$$R^2 = 0.9983$$

لاحظ أن: $\hat{Y}_i^2$ و$\hat{Y}_i^3$ في (9.4.13) تم الحصول عليهما من (8.4.13) والآن بتطبيق اختبار F نحصل على:

$$F = \frac{(0.9983 - 0.8409)/2}{(1 - 0.9983)/(10 - 4)} = 284.4035 \qquad (10.4.13)$$

يمكن للقارئ بسهولة أن يثبت أن هذه القيمة للـF لها معنوية إحصائية عالية، مما يعني أن النموذج (8.4.13) نموذج غير صحيح. بالطبع سنصل إلى نفس الاستنتاج مستخدمين الاختبار البصري للبواقي أو قيمة إحصاء d لـ Durbin-Watson.

إحدى مميزات استخدام RESET هي سهولة التطبيق، حيث إنها لا تتطلب تحديد النموذج البديل. ولكنها في نفس الوقت تعتبر نقطة ضعف في هذه الطريقة، حيث إن معرفة النموذج غير صحيح فقط لا يساعدنا في اختيار بديل أفضل.

اختبار مضروب lagrange (LM) لإضافة المتغيرات:
Lagrange Multiplier (LM) Test for Adding Variables

هذا الاختبار يعتبر بديلاً لاختبار RESET لـ Ramsey. لتوضيح هذا الاختبار، سنستعرضه وفقًا لمثالنا السابق التوضيحي.

إذا قارنا بين دالة التكلفة الخطية (6.4.13) ودالة التكلفة التكعيبية (4.4.13)، الأولى تعتبر صورة مقيدة من الأخيرة (تذكر ما سبق وناقشناه بخصوص المربعات الصغرى المقيدة في الفصل 8).

الانحدار المقيد (6.4.13) يفترض أن معامل الحدود المربعة والمكعبة من الناتج تساوي الصفر. لاختبار ذلك باستخدام أسلوب LM نقوم بالخطوات التالية:

1 - قدر الانحدار المقيد (6.4.13) باستخدام OLS واحصل على البواقي $\hat{u}_i$.

2 - اذا كان في الحقيقة الانحدار المقيد (4.4.13) هو النموذج الصحيح، فإن البواقي التي حصلنا عليها من (6.4.13) ستكون مرتبطة مع الحدود المربعة و المكعبة للناتج أي X_i^2 و X_i^3.

3 - وذلك يعني أن نقوم بعمل انحدار لـ$\hat{u}_i$ التي حصلنا عليها من الخطوة 1 على كل المتغيرات المنحدرة (مشتملة على المتغيرات الموجودة في النموذج المقيد) وذلك في مثالنا الحالي.

يعني التالي :

$$\hat{u}_i = \alpha_1 + \alpha_2 X_i + \alpha_3 X_i^2 + \alpha_4 X_i^3 + v_i \qquad (11.4.13)$$

حيث v هو حد الخطأ ومتوافر فيه الخصائص التقليدية المعروفة .

4 - في حالة إحكام العينات الكبيرة ، أثبت Engle أن n (حجم العينة) مضروب في قيمة R^2 المقدرة من الانحدار (المساعد) (11.4.13) تتبع توزيع كلي التربيعي بدرجة حرية تساوي عدد القيود المفروضة في نموذج الانحدار المقيد ، وتلك الدرجة تساوي أثنين في مثالنا الحالي ، حيث إن القيمة X_i^2 و X_i^3 تم إسقاطهما من النموذج .[25]

رمزيًا لدينا التالي :

$$nR^2 \underset{\text{asy}}{\sim} \chi^2_{(\text{number of restrictions})} \qquad (12.4.13)$$

حيث إن asy تعني تقاربيًا ، أي في الحالات التي يكون فيها حجم العينة كبيراً .

5 - إذا زادت قيمة chi-square التي حصلت عليها من (12.4.13) من القيمة الحرجة عند مستوى المعنوية المختار ، نرفض الانحدار المقيد ، وبخلاف ذلك نقبله .

نتائج الانحدار في مثالنا الحالي كالتالي :

$$\hat{Y}_i = 166.467 + 19.333 X_i \qquad (13.4.13)$$

حيث Y هي التكلفة الكلية و X هي الناتج . الأخطاء القياسية لهذا الانحدار معطاة في جدول (1.13) .

إذا قمنا بعمل انحدار لبواقي (13.4.13) كما هو مقترح في الخطوة 3 . نحصل على النتائج التالية :

$$\widehat{u}_i = -24.7 + 43.5443 X_i - 12.9615 X_i^2 + 0.9396 X_i^3$$
$$\text{se} = (6.375) \quad (4.779) \quad (0.986) \quad (0.059) \qquad (14.4.13)$$
$$R^2 = 0.9896$$

على الرغم من أن حجم العينة يساوي 10 ، وهو ليس بالحجم الكبير . إلا أننا نرغب في توضيح آلية LM كالتالي سنحصل على $nR^2 = (10)\ (0.9896) = 9.896$ عن

(25) R. F. Engle, "A General Approach to Laggrangian Multiplier Model Diagnostics," Journal of Econometrics, vol. 20, 1982, pp. 83-104.

جدول كاي التربيعي، وباستخدام 2 درجات حرية، و 1% مستوى معنوية، نجد أن قيمة كاي التربيعية تساوي تقريبًا 9.21. وبالتالي القيمة المشاهدة والمساوية لـ 9.896 لها معنوية إحصائية عند مستوى المعنوية 1% ونستنتج أنه يجب رفض النموذج المقيد (أي دالة التكلفة الخطية). ونلاحظ أننا قد توصلنا إلى نفس النتيجة باستخدام اختبار RESET لـ Ramsey.

5.13 أخطاء القياس: ERRORS OF MEASUREMENT

في كل ما سبق، تم الافتراض صراحةً بأن المتغير التابع Y، والمتغيرات المفسرة X's تم قياسها بدون أخطاء. وبالتالي في انحدار نفقات الاستهلاك على الدخل وثروة رب الأسرة، افترضنا أن بيانات هذه المتغيرات "دقيقة"، وبالتالي لا مجال للتخمين أو التدثير أو التقريب بأي شكل منتظم مثل التقريب إلى مئة دولار وهكذا.

للأسف هذا الوضع النموذجي لا يحدث كثيرًا في الواقع العملي، وذلك وفقًا للعديد من الأسباب مثل أخطاء عدم الاستجابة، وأخطاء التسجيل، وأخطاء الحساب. وبغض النظر عن السبب فإن أخطاء القياس مشكلة محتملة عن التطبيق العملي، وهي تعتبر أحد مصادر تحيز التوصيف. وسنستعرض فيما يلي العواقب المحتملة لذلك:

أخطاء قياس في المتغير التابع Y:

Errors of Measurement in the Dependent Variable Y

اعتبر النموذج التالي:

$$Y_i^* = \alpha + \beta X_i + u_i \quad (1.5.13)$$

حيث Y_i^* = نفقات الاستهلاك الدائمة[26]

X_i = الدخل الحالي

u_i = خطأ عشوائي

بما أن Y_i^* لا يمكن قياسها مباشرة، تستخدم النفقات المشاهدة (المتغير Y_i) كالتالي:

$$Y_i = Y_i^* + \varepsilon_i \quad (2.5.13)$$

حيث ε_i ترمز للخطأ في قياس Y_i^*. وبالتالي بدلاً من تقدير (1.5.13)، فإننا نقدر:

(26) هذه العبارة تعود إلى Milton Friedman انظر أيضًا تدريب (8.13).

$$\begin{aligned} Y_i &= (\alpha + \beta X_i + u_i) + \varepsilon_i \\ &= \alpha + \beta X_i + (u_i + \varepsilon_i) \\ &= \alpha + \beta X_i + v_i \end{aligned} \tag{3.5.13}$$

حيث $v_i = u_i + \varepsilon_i$ وهو يعتبر حد خطأ مركب، يشتمل على خطأ المجتمع (والذي يمكن تسميته معادلة حد الخطأ) وحد الخطأ في القياس.

للتبسيط افترض أن $E(u_i) - E(\varepsilon_i) = 0$ ، $\text{cov}(X_i, u_i) = 0$ (والذي يعتبر فرضًا من فروض الانحدار الخطي التقليدي) و $\text{cov}(X_i, \varepsilon_i) = 0$ أي أن أخطاء القياس ف Y_i^* غير مرتبطة مع X_i و $\text{cov}(u_i, \varepsilon_i) = 0$ أي أن خطأ المعادلة وخطأ القياس غير مرتبطين. بالإضافة إلى بعض الفروض الخاصة بـβ المقدرة سواء من (1.5.13) أو (3.5.13) والتي تنص على أنها مقدر غير متحيز للمعلمة الحقيقية β (انظر تمرين 7.13). وبالتالي أخطاء القياس الموجودة في المتغير التابع Y لا تؤثر على خاصية عدم التحيز لمقدرات OLS. عمومًا التباين والأخطاء القياسية للـβ المقدرة من (1.5.13) و (3.5.13) ستكون مختلفة، باستخدام المعادلات المعتادة (انظر الفصل 3)، نحصل على:

$$\text{نموذج (1.5.13)} : \quad \text{var}(\hat{\beta}) = \frac{\sigma_u^2}{\sum x_i^2} \tag{4.5.13}$$

$$\text{نموذج (3.5.13)} : \quad \text{var}(\hat{\beta}) = \frac{\sigma_v^2}{\sum x_i^2} = \frac{\sigma_u^2 + \sigma_\varepsilon^2}{\sum x_i^2} \tag{5.5.13}$$

نلاحظ أن التباين الأخير أكبر من التباين الأول.[27] وبالتالي على الرغم من وجود أخطاء القياس في المتغير التابع Y فإن مقدرات المعالم والتباينات مازالت غير متحيزة، إلا أن التباين المقدر أكبر من نظيره في الحالة التي لا يوجد فيها مثل هذا الخطأ من أخطاء القياس.

أخطاء القياس في المتغير المفسر X:

Errors of Measurement in the Explanatory Variable *X*

الآن افترض أنه بدلاً من (1.5.13) لدينا النموذج التالي:

(27) لكن نلاحظ أن هذا التباين مازال غير متحيز، حيث إنه وفقًا للشروط الموضوعة الخطأ المركب $v_i = u_i + \varepsilon_i$ مازال مستوفيًا لكل الفروض الخاصة بطريقة المربعات الصغرى.

$$Y_i = \alpha + \beta X_i^* + u_i \qquad (6.5.13)$$

حيث Y_i = نفقات الاستهلاك الحالية

X_i^* = الدخل الدائم

u_i = حد الخطأ العشوائي (معادلة الخطأ)

وافترض أنه بدلاً من مشاهدة X_i^*، نشاهد التالي:

$$X_i = X_i^* + w_i \qquad (7.5.13)$$

حيث w_i تمثل خطأ في قياس X_i^*. وبالتالي بدلاً من تقدير (6.5.13) نقدر التالي:

$$\begin{aligned} Y_i &= \alpha + \beta(X_i - w_i) + u_i \\ &= \alpha + \beta X_i + (u_i - \beta w_i) \qquad (8.5.13) \\ &= \alpha + \beta X_i + z_i \end{aligned}$$

حيث $z_i = u_i - \beta w_i$ ، وبالتالي لدينا مزيج من معادلة الأخطاء وأخطاء القياس.

والآن حتى لو افترضنا أن w_i لها وسط يساوي الصفر، وغير مرتبطة تسلسليًا، وغير مرتبطة بالـ u_i، لا نستطيع افتراض أن حد الخطأ المركب z_i مستقل عن المتغير المفسر X_i، وذلك بسبب التالي [افترض أن $E(z_i) = 0$]

$$\begin{aligned} \text{cov}(z_i, X_i) &= E[z_i - E(z_i)][X_i - E(X_i)] \\ &= E(u_i - \beta w_i)(w_i) \qquad \text{using } (13.5.7) \\ &= E(-\beta w_i^2) \qquad (9.5.13) \\ &= -\beta\sigma_w^2 \end{aligned}$$

وبالتالي المتغير المفسر وحد الخطأ في (8.5.13) مرتبطان، وذلك يخالف فرضًا مهمًا في نموذج الانحدار الخطي التقليدي، وهو الفرض القائل بأن المتغير المفسر غير مرتبط مع حد الخطأ العشوائي. إذا خالفنا هذا الفرض، فإنه يمكن إثبات أن مقدرات OLS لم تعد فقط متحيزة، ولكن غير متسقة أيضًا، أي ستكون مقدرات متحيزة حتى لو زاد حجم العينة n وأصبح غير محدود.[28]

---

(28) كما نرى من الملحق A، $\hat{\beta}$ مقدر متسق لـ β اذا حدث التالي: كلما زاد حجم العينة n كلما اقترب توزيع المعاينة لـ $\hat{\beta}$ من القيمة الحقيقية لـ β. وبشكل فني يمكن التعبير عن العبارة السابقة كالتالي $\text{plim}_{n\to\infty}\hat{\beta} = \beta$. وكلما نلاحظ في ملحق A. الاتساق خاصية مرتبطة بأحجام العينات الكبيرة وعادة ما يستخدم لدراسة مقدر لا يمكن تحديد خصائصه في حالة العينات الصغيرة أو المحدودة (مثل عدم التحيز).

بالنسبة للنموذج (8.5.13) مثبت في الملحق A13، الفقرة 3A.13 أن :

$$\text{plim}\,\hat{\beta} = \beta\left[\frac{1}{1+\sigma_w^2/\sigma_{X^*}^2}\right] \quad (10.5.13)$$

حيث σ_w^2 و $\sigma_{X^*}^2$ تمثل تباين w_i و X^* بالترتيب و $\hat{\beta}$ plim تعني النهاية الاحتمالية لـ β.

بما أن المقدار ما بين الأقواس متوقع أن يكون أقل من 1 (لماذا؟) فإن (10.5.13) توضح أنه حتى إذا زاد حجم العينة بشكل غير محدود فإن $\hat{\beta}$ لن يؤول إلى β. في واقع الأمر إذا افترضنا أن β موجبة، فإن $\hat{\beta}$ ستقدر قيمة β بأقل من قيمتها الحقيقية، وبالتالي يكون التحيز أقل من الصفر. بالطبع إذا لم يكن هناك أي أخطاء في القياس في X (أي أن $\sigma_w^2 = 0$) فإن $\hat{\beta}$ سيعتبر مقدراً متسقًا بالنسبة للـ β.

وبالتالي تمثل أخطاء القياس مشكلة حقيقية عندما تظهر في المتغير أو المتغيرات المفسرة، حيث إنهم يؤدون إلى عدم إمكانية الحصول على مقدر متسق. بالطبع، كما سبق ورأينا، إذا ظهرت أخطاء القياس في المتغير التابع فقط، فإن المقدرات لا تزال غير متحيزة ومتسقة أيضًا. أما إذا ظهرت أخطاء القياس في المتغيرات المفسرة. ما الحل؟ الإجابة ليست بالسهولة المتوقعة. إحدى الإجابات المتطرفة، أننا يمكن أن نفترض أن σ_w^2 صغير للغاية مقارنةً مع $\sigma_{X^*}^2$، وبالتالي لكل التطبيقات العلمية يمكن أن نفترض عدم حدوث المشكلة، ونستمر في استخدام تقدير OLS التقليدي. بالطبع، هنا يمكننا مشاهدة أو قياس أي من σ_w^2 أو $\sigma_{X^*}^2$ وبالتالي لا يمكن الحكم على مدى تأثيرهما.

إحدي الإجابات الأخرى والتي يمكن اعتبارها أسلوب علاج للمشكلة الحالية هو التعامل مع المتغيرات المساعدة، وعلى الرغم من أنها ستكون مرتبطة بدرجة كبيرة مع المتغيرات X الأصلية، إلا أنها غير مرتبطة مع حد خطأ المعادلة أو حدود أخطاء القياس (أي u_i و w_i). إذا استطاع إيجاد مثل هذا المتغير المساعد، بالتالي يمكنك الحصول على مقدر متسق للـ β.

ولكن هذه العملية، التحدث عنها أسهل بكثير من تنفيذها. في الواقع العملي، ليس من السهل إيجاد مثل هذه المتغيرات المساعدة، فنحن في الأغلب كثيرو الشكوى من المناخ، ولكننا لا تستطيع فعل أي شيء حياله. وإلى جانب صعوبة الحصول على المتغيرات المساعدة، فإنه أيضًا من الصعب معرفة ما إذا كانت المتغيرات المساعدة بالفعل مستقلة عن حدود الخطأ u_i و w_i أم لا.

في الأدبيات هناك اقتراحات أخرى لحل هذه المشكلة.[29]

ولكن أغلب هذه الحلول خاصة بأوضاع محددة ومبنية على فروض مقيدة عديدة. وبالتالي لا توجد حلول محددة لمشكلة أخطاء القياس. ولذلك لابد من القياس الدقيق على قدر المستطاع للبيانات محل الدراسة.

مثال An Example

سنلخص الفقرة السابقة في المثال التالي.

جدول (2.13) يعطي بيانات افتراضية عن نفقات الاستهلاك الحقيقية Y^*، الدخل الحقيقي X^*، الاستهلاك المقاس Y والدخل المقاس X، الجدول أيضاً يتضح فيه كيفية قياس تلك المتغيرات.[30]

أخطاء القياس في المتغير التابع Y فقط

Measurement Errors in the Dependent Variable Y only

بناء على البيانات المعطاة، دالة الاستهلاك الحقيقية هي

$$\begin{aligned} \hat{Y}_i^* &= 25.00 + 0.6000X_i^* \\ &\quad (10.477) \quad (0.0584) \\ t &= (2.3861) \quad (10.276) \\ R^2 &= 0.9296 \end{aligned} \tag{11.5.13}$$

في حين أنه إذا استخدمنا Y_i بدلاً من Y_i^* سنحصل على:

$$\begin{aligned} \hat{Y}_i &= 25.00 + 0.6000X_i^* \\ &\quad (12.218) \quad (0.0681) \\ t &= (2.0461) \quad (8.8118) \\ R^2 &= 0.9066 \end{aligned} \tag{12.5.13}$$

كما يتضح من النتائج السابقة وبناء على النظرية، القيمة المقدرة للمعاملات تظل كما هي. الأثر الوحيد لخطأ القياس في المتغير التابع هو أن الأخطاء القياسية المقدرة للمعاملات تزداد [انظر (5.5.13)] ويتضح ذلك في (12.5.13) عموماً لاحظ أن معاملات الانحدار في (11.5.13) و (12.5.13) متساوية، حيث إن البيانات تم توليدها لتتماشى مع فروض نموذج أخطاء القياس.

أخطاء القياس في X:

Errors of Measurement in X

نحن نعرف أن الانحدار الحقيقي هو (11.5.13)، افترض الآن أنه بدلاً من استخدام X_i^*، استخدمنا X_i (لاحظ أنه في الواقع نادراً ما تستطيع قياس X_i^*). نتائج الانحدار ستكون كالتالي:

$$\begin{aligned} \hat{Y}_i^* &= 25.992 + 0.5942X_i \\ &\quad (11.0810) \quad (0.0617) \\ t &= (2.3457) \quad (9.6270) \\ R^2 &= 0.9205 \end{aligned} \tag{13.5.13}$$

هذه النتائج عندما تكون هناك أخطاء قياس في المتغير أو المتغيرات المفسرة.

(29) انظر Thomas B. Fomby, R. Carter Hill, and Stanley R. Johnson, Advanced Econometric Methods, Springer-Verlag, New York, 1984, pp. 273-277. مزيد من التفاصيل عن الانحدار والمتغيرات المساعدة.

(30) خالص شكرى لـ Keneth J. White لتصميم ذلك المثال. انظر في SHAZAM, for use wit Damodar Gujarati, Basic Econometrics, Septmber 1985, pp. 117-121.

المعاملات المقدرة متحيزة. لحسن الحظ في هذا المثال التحيز صغير نسبيًا – من (10.5.13) نرى أن التحيز يعتمد على $\sigma_w^2 / \sigma_{X^*}^2$، وعند توليد البيانات افترضنا أن $\sigma_w^2 = 36$ و $\sigma_{X^*}^2 = 3667$ مما يجعل معامل التحيز صغيراً يساوي تقريباً 0.98% (36/3667).

سنترك للقارئ تحديد ما يحدث إذا كان هناك خطأ في قياس كل من Y و X.

أي عندما تقوم بعمل انحدار لـ Y_i على X_i بدلاً من Y^*_i على X^*_i (انظر تمرين 23.13).

جدول (2.13) بيانات افتراضية عن Y^* (نفقات الاستهلاك الحقيقية)، X^* (الدخل الحقيقي) و Y (نفقات الاستهلاك المقاسة) و X (الدخل المقاس) البيانات كلها بالدولار

Y^*	X^*	Y	X	ε	w	u
75.4666	80.00	67.6011	80.0940	−7.8655	0.0940	2.4666
74.9801	100.00	75.4438	91.5721	0.4636	−8.4279	−10.0199
102.8242	120.00	109.6956	112.1406	6.8714	2.1406	5.8242
125.7651	140.00	129.4159	145.5969	3.6509	5.5969	16.7651
106.5035	160.00	104.2388	168.5579	−2.2647	8.5579	−14.4965
131.4318	180.00	125.8319	171.4793	−5.5999	−8.5207	−1.5682
149.3693	200.00	153.9926	203.5366	4.6233	3.5366	4.3693
143.8628	220.00	152.9208	222.8533	9.0579	2.8533	−13.1372
177.5218	240.00	176.3344	232.9879	−1.1874	−7.0120	8.5218
182.2748	260.00	174.5252	261.1813	−7.7496	1.1813	1.2748

Note: The data on X^* are assumed to be given. In deriving the other variables the assumptions made were as follows: (1) $E(u_i) = E(\varepsilon_i) = E(w_i) = 0$; (2) $\text{cov}(X, u) = \text{cov}(X, \varepsilon) = \text{cov}(u, \varepsilon) = \text{cov}(w, u) = \text{cov}(\varepsilon, w) = 0$; (3) $\sigma_u^2 = 100$, $\sigma_\varepsilon^2 = 36$, and $\sigma_w^2 = 36$; and (4) $Y_i^* = 25 + 0.6X_i^* + u_i$, $Y_i = Y_i^* + \varepsilon_i$, and $X_i = X_i^* + w_i$.

6.13 التوصيف الخاطئ لحد الخطأ العشوائي :

INCORECT SPECIFICATION OF THE STOCHASTIC ERROR TERM

المشكلة الأكثر ظهوراً أمام الباحث، هي توصيف حد الخطأ u_i داخل نموذج الانحدار. فيما أن حد الخطأ لا يشاهد مباشرةً، فإنه لا توجد طريقة لتحديد الشكل الذي يظهر به في النموذج. لتوضيح ذلك دعنا نعود إلى النماذج المعطاة في (8.2.13) و (9.2.13). للتبسيط، افترضنا عدم وجود الحد الثابت المقطوع من المحور الصادي في النموذج، وافترضنا أيضًا أن u_i الموجود في (8.2.13) يستوفى فيه $\ln u_i$ كل فروض OLS التقليدية.

إذا افترضنا أن (8.2.13) هو النموذج " الصحيح "، ولكن قدرنا (9.2.13) ما هي عواقب ذلك؟ يتضح من ملحق A13، الفقرة 4A.13 انه إذا كان $\ln u_i \sim N(0, \sigma^2)$ فإن

$$u_i \sim \log\text{normal}\left[e^{\sigma^2/2},\ e^{\sigma^2}\left(e^{\sigma^2} - 1\right)\right] \qquad (1.6.13)$$

وكنتيجة لذلك يكون لدينا التالي:

$$E(\hat{\alpha}) = \beta e^{\sigma^2/2} \qquad (2.6.13)$$

حيث e هي أساس اللوغاريتم الطبيعي.

كما ترى، $\hat{\alpha}$ مقدر متحيز، حيث إن قيمة المتوسط لا تساوي القيمة الحقيقية لـ β.

سنستعرض المزيد من التفاصيل عن توصيف حد الخطأ العشوائي في الفصل الخاص بنماذج الانحدار غير الخطية في المعالم.

7.13 النماذج المتداخلة في مقابل غير المتداخلة : NESTED VERSUS NON-NESTED MODELS

عند تطبيق اختبارات التوصيف، لابد من التفرقة بين النماذج المتداخلة والنماذج غير المتداخلة. للقيام بذلك، اعتبر النماذج التالية:

نموذج A : $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \beta_5 X_{5i} + u_i$$

نموذج B : $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

النموذج B يسمى نموذج متداخل مع النموذج A، حيث إنه يمثل حالة خاصة من النموذج A. فإذا قدرنا النموذج A، واختبرنا الفرض الخاص بأن $\beta_4 = \beta_5 = 0$ وإذا لم نرفض مثل هذا الفرض مثلاً بناء على اختبار F.[31] فإن نموذج A يصبح هو النموذج B إذا اضفنا المتغير X_4 إلى النموذج B، فإن نموذج A سيصبح هو النموذج β إذا كانت $\beta_5 = 0$، وهنا سنستخدم اختبار t لاختبار فرض تساوي معامل X_5 للصفر.

ويدون استخدم العبارة التالية للتوصيف، فإن اختبارات أخطاء التوصيف التي سبق مناقشتها، واختبار F المقيد السابق مناقشته في الفصل (8)، يعتبران اختبارات مهمة لفرض التداخل.

والآن دعنا نعتبر النماذج التالية:

نموذج C : $$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i$$

نموذج D : $$Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + v_i$$

حيث إن X's وZ's مختلفان، نقول ان النماذج C و D نماذج غير متداخلة، حيث إنه لا يمكن اشتقاق أي نموذج منهما من الآخر، ولا يمكن اعتبار أي منهما حالة خاصة من الآخر. في الاقتصاد، كما في العلوم الأخرى، قد توجد أكثر من نظرية لشرح

(31) بشكل أكثر عمومية، يمكن اختيار اختبار نسبة الإمكان أو اختيار Wald أو اختبار المضروب لـ Lagrange الذي سبق مناقشته باختصار في الفصل 8.

ظاهرة واحدة. فمثلاً الماليين قد يؤكدون على دور المال في شرح التباين والاختلاف في GDP، أما الكينزيين فيؤكدون على دور نفقات الحكومة.

ونلاحظ هنا أنه يمكن السماح بوجود متغيرات منحدرة مشتركة بين النماذج C و D. فمثلاً، X_3 يمكن أن يتضمن في النموذج D و Z_2 يمكن إدخاله في النموذج C. وستظل أيضًا نماذج غير متداخلة، حيث إن النموذج C لا يحتوي على Z_3، والنموذج D لا يحتوي على X_2.

حتى إذا كانت المتغيرات واحدة، فإن الشكل الدالي قد يجعل النموذجين غير متداخلين. مثلاً اعتبر النموذج التالي:

النموذج E : $$Y_i = \beta_1 + \beta_2 \ln Z_{2i} + \beta_3 \ln Z_{3i} + w_i$$

النماذج D و E غير متداخلة. فلا يمكن اعتبار أي من النموذجين حالة خاصة في النموذج الآخر.

وحيث إننا سبق وتناولنا اختبارات النماذج المتداخلة (اختبارات t و F)، فإننا في الفقرة التالية، سنناقش بعض الاختبارات الخاصة بالنماذج غير المتداخلة، والتي سبق وأطلقنا عليها أخطاء التوصيف الخاطئ.

8.13 اختبارات فروض عدم التداخل:
TESTS OF NON-NESTED HYPOTHESES

بناء على Harvery [32]، هناك أسلوبان لاختبار فروض عدم التداخل: (1) أسلوب التمييز، فيكون معطى لدى الباحث نموذجان أو أكثر متنافسين، ويختار نموذجًا بناء على جودة التوفيق و (2) أسلوب المميز (تسمية خاص بي)، حيث عند البحث عن نموذج نأخذ في الاعتبار المعلومات المتاحة من نماذج أخرى.

سنستعرض هذه الأساليب باختصار في الفقرة التالية:

أسلوب التمييز: The Discrimination Approach

اعتبر النماذج C و D المعطاة أعلى. بما أن كلاً من النموذجين يشتمل على نفس المتغير التابع. يمكن الاختبار بين النموذجين بناء على أساليب جودة التوفيق مثل R^2

(32) Andrew Harvey, The Econoetric Analysis of Time Series, 2d ed., The MIT Press, Cambridge, Mass., 1990, Chap. 5.

أو R^2 المعدلة، والتي سبق واستعرضناها. ولكن ضع في الاعتبار أنه عند مقارنة نموذجين أو أكثر، يجب أن يكون لدينا نفس المتغير المنحدر عليه، وبجانب هذه المعايير، هناك معايير أخرى يمكن استخدامها مثل Akaile's information criterion (AIC) و Schwarz's information criterion (SIC) و mallows's C_p criterion.

قد سبق وناقشنا هذه المعايير في الفقرة (9.13)، أغلب الحزم الإحصائية الحديثة لدينا واحد أوأكثر من هذه المعايير يمكن للباحث أن يستخدمها في إطار نماذج الانحدار. في الفقرة الأخيرة من هذا الفصل، سنستعرض هذه المعايير باستخدام مثال توضيحي. وبناء على معيار أو أكثر من هذه المعايير، يتم اختيار نموذج له $\bar{R}^2$ عالية أو أقل AIC أو SIC وهكذا.

أسلوب المميز: The discerning Approach

اختبار *F* غير المتداخل أو الشامل: The Non-Nested *F* Test or Encompassing *F* Test

اعتبر النماذج C و D السابق تقديمهما. كيف يمكنك الاختيار بين هذين النموذجين؟ لهذا الفرض افترض أننا نقدر النموذج المتداخل التالي:

نموذج F: $$Y_i = \lambda_1 + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \lambda_4 Z_{2i} + \lambda_5 Z_{3i} + u_i$$

لاحظ أن نموذج F متداخل أو شامل على نموذج C و D. ولكن لاحظ أن C غير متداخل مع D و D غير متداخل أيضًا مع C، وبالتالي فهما نماذج غير متداخلة.

الآن يكون النموذج C نموذج صحيحًا إذا كان $\lambda_4 = \lambda_5 = 0$، ويكون النموذج D صحيحًا إذا كان $\lambda_2 = \lambda_3 = 0$، هذا الاختبار يمكن عمله باستخدام اختبار F التقليدي، ومن هنا جاءت تسميته باختبار *F* غير المتداخل.

عمومًا هناك بعض المشاكل المتعلقة بعملية الاختبار. أولاً، إذا كان *X*'s و *Z*'s مرتبطين بدرجة عالية، فإنه كما سبق وذكرنا في فصل التعدد الخطي، سيكون لدينا واحد أو أكثر من λ's غير معنوي احصائيًا على الرغم من أن الباحث، بناء على اختبار F، سيرفض الفرض القائل بأن معاملات الميل تساوي الصفر آنيًا. في مثل هذه الحالة، ليست لديك طريقة تستطيع بها تحديد ما إذا كان النموذج C أو النموذج D هو النموذج السليم. ثانيًا، هناك مشكلة أخرى، افترض أنك اخترت النموذج C

كالنموذج المرجعي أو الافتراضي المرجعي، ووجدت أن جميع معاملاته معنوية، وعند إضافة Z_2 أو Z_3 أو كليهما إلى النموذج ووجدت انه باستخدام اختبار F أن الاضافة الخاصة بهما في تفسير مجموع المربعات المفسرة (ESS) غير معنوي. بالتالي ستختار النموذج C.

ولكن افترض أنه بدلاً من ذلك اخترت النموذج D كالنموذج المرجعي، ووجدت أن معاملات معنوية إحصائيًا. ولكن عند إضافة X_2 أو X_3 أو كليهما إلى النموذج وجدت باستخدام اختبار F أن الإضافة الخاصة بهما في تفسير مجموع المربعات المفسرة (ESS) غير معنوي. وبالتالي هنا يكون اختيار النموذج D على أساس انه النموذج الصحيح غير دقيق. وبالتالي "اختبار النموذج المرجعي قد يحدد نتائج النموذج المختار"[33] خصوصًا إذا كان هناك ارتباط متعدد عال بين المتغيرات المنحدرة المتنافسة. وأخيرًا اختبار F المتداخل قد لا يكون له أي معنى حقيقي اقتصادي.

مثال توضيحي: نموذج LOUIS MODEL

لتحديد ما إذا كانت التغيرات في GNP الإسمي يمكن تفسيرها بالتغيرات في المعروض من المال (الماليين) أو بالتغيرات في نفقات الحكومة (الكينزيين). دعنا نعتبر النماذج التالية:

$$\dot{Y}_t = \alpha + \beta_0 \dot{M}_t + \beta_1 \dot{M}_{t-1} + \beta_2 \dot{M}_{t-2} + \beta_3 \dot{M}_{t-3} + \beta_4 \dot{M}_{t-4} + u_{1t}$$

$$= \alpha + \sum_{i=0}^{4} \beta_i \dot{M}_{t-i} + u_{1t} \qquad (1.8.13)$$

$$\dot{Y}_t = \gamma + \lambda_0 \dot{E}_t + \lambda_1 \dot{E}_{t-1} + \lambda_2 \dot{E}_{t-2} + \lambda_3 \dot{E}_{t-3} + \lambda_4 \dot{E}_{t-4} + u_{2t}$$

$$= \gamma + \sum_{i=0}^{4} \lambda_i \dot{E}_{t-i} + u_{2t} \qquad (2.8.13)$$

حيث $\dot{Y}_t$ = معدل النمو في GNP الإسمي عند الزمن t

$\dot{M}_t$ = معدل النمو في المعروض من المال (نسخة M_1) عند الزمن t

$\dot{E}_t$ = معدل النمو في نفقات الحكومة الوظيفية عند الزمن t

عمومًا لاحظ أن (1.8.13) و (2.8.13) أمثلة للنماذج ذات الفترات الزمنية المتأخرة، وهذا الموضوع ناقشناه في الفصل (17). أما الآن فلاحظ أن أثر التغيير في المعروض من المال أو نفقات الحكومة على GDP موزعة على فترة زمنية ولا يتم التعامل معه لحظيًا.

(33) Thomas B. Fomby, R. Carter Hill, and Stanley r. Johnson, Advanced Econometric Methods, Springer Verlag, New York, 1984, p. 416.

وحيث إنه من الصعب مسبقًا الحكم على النموذجين السابقين المتنافسين . دعنا نقوم بالتالي :

$$\dot{Y}_t = \text{constant} + \sum_{i=0}^{4}\beta_i \dot{M}_{t-i} + \sum_{i=0}^{4}\lambda_i \dot{E}_{t-i} + u_{3t} \qquad (3.8.13)$$

هذا النموذج المتداخل هو نموذج شهير الاستخدام يسمى نموذج St-louis وهو مؤيد للنظرية المالية . وعند تقدير هذا النموذج حصلنا على النتائج التالية خلال الفترة 1953-I إلى 1976-IV وذلك في الولايات المتحدة الأمريكية وكانت النتائج كالتالي (نسب t موضحة بين الأقواس) .[34]

Coefficient	Estimate	Coefficient	Estimate	
β_0	0.40 (2.96)	λ_0	0.08 (2.26)	
β_1	0.41 (5.26)	λ_1	0.06 (2.52)	
β_2	0.25 (2.14)	λ_2	0.00 (0.02)	
β_3	0.06 (0.71)	λ_3	−0.06 (−2.20)	(13.8.4)
β_4	−0.05 (−0.37)	λ_4	−0.07 (−1.83)	
$\sum_{i=0}^{4}\beta_i$	1.06 (5.59)	$\sum_{i=0}^{4}\lambda_i$	0.03 (0.40)	
				$R^2 = 0.40$
				$d = 1.78$

ما الذي يمكن استنتاجه من هذه النتائج حول تفضيل نموذج على الآخر؟ إذا اعتبرنا الأثر التجميعي للتغير الوحدة في $\dot{M}$ أو $\dot{E}$ على $\dot{Y}$ ، نحصل على التالي ، بالترتيب، $\sum_{i=0}^{4}\beta_i = 1.06$ و $\sum_{i=0}^{4}\lambda_i = 0.03$ والأول له معنوية إحصائية ،في حين أن الأخير ليس له معنوية إحصائية . وهذه المقارنة ستؤدي إلى تفضيل الفرض المالي أي أن التغيير في المعروض من المال يمكنه أن يحدد التغييرات في GNP (الإسمي) . ومتروك للقارئ كتمرين أن يقيم هذا الفرض بشكل دقيق .

اختبار Davidson-MacKinnon J Test[35]

نظرًا للمشاكل السابق عرضها والخاصة بعملية اختبار F غير المتداخل ، هناك بديل آخر مقترح . هذا البديل هو اختبار Davidson-MacKinnon J . لشرح هذا الاختبار ، افترض أننا نريد المقارنة بين النموذج C والنموذج D . طريقة اختبار J تعتمد على الخطوات التالية :

1 – قدر النموذج D ومنه احصل على قيم Y المقدرة، $\hat{Y}_i^D$.

(34) انظر في Keith M. Carlson, "Does the St. Louis Equation Now Believe in Fiscal Policy?" Review, Federal Reserve Bank of St. Louis, vol. 60, no. 2, February 1978, p. 17, table IV.

(35) R. Davidson and J. G. MacKinnon, "Several Tests for Model Specification in the Presence of Alternative Hypothesses, "Econometrica, vol. 49, 1981, pp. 781-793.

2 - أضف قيم Y المتنبأ بها من الخطوة 1 كمتغير منحدر جديد إلى النموذج C، وقدر النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 \hat{Y}_i^D + u_i \qquad (5.8.13)$$

حيث $\hat{Y}_i^D$ هي التي حصلنا عليها من الخطوة 1. هذا النموذج يعتبر مثالاً لمبدأ الشمولية كما في منهجية Hendry.

3 - باستخدام اختبار t، اختبر الفرض $\alpha_4 = 0$.

4 - إذا لم يتم رفض الفرض الخاص بأن $\alpha_4 = 0$، فإنه يمكن قبول (أي عدم رفض) النموذج C كالنموذج الصحيح، حيث $\hat{Y}_i^D$ الموجودة في (5.8.13) والتي تمثل أثر المتغيرات وغير الموجودة في النموذج C ليس لها إضافة تفسيرية تفوق الموجودة في النموذج C. بمعنى آخر، النموذج C يشتمل على النموذج D بمعنى أن الأخير لا يحتوي على أي معلومات إضافية تحسن من أداء النموذج C. وبنفس المنطق، اذا تم رفض الفرض العدمي، فإن النموذج C لا يمكن اعتباره النموذج الصحيح (لماذا؟).

5 - والآن دعنا نبدل الفروض أو نماذج C و D. فدعنا الآن نقدر النموذج C أولاً ونقدر قيم Y من هذا النموذج، ثم نضيفها كمتغير منحدر إلى (5.8.13)، نكرر الخطوة 4 ونقرر قبول أو رفض النموذج D في مقابل النموذج C. وبشكل أكثر تحديداً، نقدر النموذج التالي:

$$Y_i = \beta_1 + \beta_2 Z_{2i} + \beta_3 Z_{3i} + \beta_4 \hat{Y}_i^C + u_i \qquad (6.8.13)$$

حيث $\hat{Y}_i^C$ تمثل قيم Y المقدرة من النموذج C. ونختبر الآن الفرض القائل بأن $\beta_4 = 0$. إذا لم نرفض هذا الفرض، نختار النموذج D أفضل من C. أما إذا تم رفض الفرض $\beta_4 = 0$ نختار C بدلاً من D، حيث إن الأخير لا يقدم أي تحسن في أداء النموذج C.

ويجب الذكر بأن اختبار J يعاني من بعض المشاكل. حيث إن الاختبارات المعطاة في (5.8.13) و (6.8.13) تتم بشكل مستقل، لدينا النتائج التالية:

اختيار: $\beta_4 = 0$

اختيار : $\beta_4 = 0$	لا نرفض الفرض	نرفض الفرض
لا نرفض الفرض	نقبل كلاً من C و D	نقبل D، ونرفض C
نرفض الفرض	نقبل C ونرفض D	نرفض كلاً من C و D

كما يتضح من الجدول السابق، لن تستطيع اختيار واحد من النموذجين إذا كانت نتيجة J تؤدي إلى قبول أو رفض النموذجين معاً. ففي حالة رفض كلاً من النموذجين، لن يساعدنا أي منهما في تفسير Y. وبنفس الطريقة إذا قبلنا النموذجين معًا، فكما لاحظ Kmenta، "البيانات ليست غنية أو كافية للتفريق بين النموذجين المفترضين". [36]

مشكلة أخرى متعلقة باختبار J أنه عند استخدام إحصاء t لاختبار معنوية المتغير Y المقدر من النماذج (5.8.13) و (6.8.13) فإحصاء t يتبع التوزيع الطبيعي القياسي تقاربيًا.

أي عند استخدام عينات ذات أحجام كبيرة نسبيًا. وبالتالي اختبار J قد لا يكون له القوة المعروفة (قوة بالمعنى الإحصائي) في العينات صغيرة الحجم، حيث سيؤدي إلى رفض الفرض العدمي الصحيح أو النماذج الصحيحة بشكل أكثر من المفروض.

مثال توضيحي : AN ILLUSTRATIVE EXAMPLE

لتوضيح اختبار J، اعتبر البيانات المعطاة في جدول (3.13). هذا الجدول يعطي بيانات عن نفقات الاستهلاك الشخصية (PPCE) والدخل الشخصي (PDPI) كلاهما مقاس بناء على 1987 دولارًا، في الولايات المتحدة الأمريكية خلال الفترة 1970 – 1991.

الآن اعتبر النماذج التالية:

(7.8.13) نموذج A :
$$PPCE_t = \alpha_1 + \alpha_2 PDPI_t + \alpha_3 PDPI_{t-1} + u_t$$

(8.8.13) نموذج A :
$$PPCE_t = \beta_1 + \beta_2 PDPI_t + \beta_3 PPCE_{t-1} + u_t$$

نموذج A يعني أن PPCE يعتمد على PDPI في الفترة الحالية والسابقة. هذا النموذج يعتبر مثالاً لنماذج الفترات الزمنية المتأخرة (انظر الفصل 17). نموذج B يفترض أن PPCE يعتمد على PDPI الحالية وعلى PPCE في فترة زمنية سابقة أيضًا. وهذا النموذج معروف باسم نماذج الانحدار الذاتي (انظر الفصل 17).

---

(36) Jan Kmenta, op. cit., p. 597

جدول (3.13) Per Capita personal consumption Expenditure (PPCE) and PER CAPITA PERSONAL DISPOSABLE INCOME (PDPI), 1987 DOLLARS. U.S., 1970-1991

Year	PPCE	PDPI	Year	PPCE	PDPI
1970	8,842	9,875	1981	10,770	12,156
1971	9,022	10,111	1982	10,782	12,146
1972	9,425	10,414	1983	11,179	12,349
1973	9,752	11,013	1984	11,617	13,029
1974	9,602	10,832	1985	12,015	13,258
1975	9,711	10,906	1986	12,336	13,552
1976	10,121	11,192	1987	12,568	13,545
1977	10,425	11,406	1988	12,903	13,890
1978	10,744	11,851	1989	13,029	14,005
1979	10,876	12,039	1990	13,044	14,068
1980	10,746	12,005	1991	12,824	13,886

السبب وراء وجود القيمة المتأخرة لـ PPCE في النموذج هو توضيح أثر الإصرار.

نتائج النماذج السابقة كانت كالتالي:

نموذج A :

$$\widehat{PPCE}_t = -1{,}299.0536 + 0.9204\ PDPI_t + 0.0931\ PDPI_{t-1}$$
$$t = \quad (-4.0378) \quad (6.0178) \quad (0.6308) \qquad (9.8.13)$$
$$R^2 = 0.9888 \qquad d = 0.8092$$

نموذج B :

$$\widehat{PPCE}_t = -841.8568 + 0.7117\ PDPI_t + 0.2954\ PPCE_{t-1}$$
$$t = \quad (-2.4137) \quad (5.4634) \quad (2.3681) \qquad (10.8.13)$$
$$R^2 = 0.9912 \qquad d = 1.0144$$

إذا كنا بصدد الاختيار بين النموذجين السابقين وفقًا لأسلوب التمييز، مثلاً باستخدام معيار R^2 العالية، سيختار الباحث (10.8.13)، بالإضافة إلى أن في (10.8.13) فقط قيم PDPI الحالية هي التي لها معنوية إحصائية (ضع في الاعتبار مشكلة التعدد الخطي).

ولكن اختيار (10.8.13) في مقابل (9.8.13) قد لا يكون سليمًا، حيث إنه للتنبأ لا يوجد فرق كبير بين قيم R^2 المقدرة.

لتطبيق اختبار J، افترض أن النموذج A هو الموجود في الفرض العدمي، أي أنه النموذج المرجعي، والنموذج B موجود في الفرض البديل. الآن وفقًا لخطوات اختبار J السابق مناقشتها نستخدم PPCE المقدرة من النموذج (10.8.13) كمتغير منحدر إضافي في النموذج A ونحصل علي النتائج التالية:

$$\widehat{PPCE}_t = 1{,}322.7958 - 0.7061 PDPI_t - 0.4357 PDPI_{t-1} + 2.1335 \widehat{PPCE}_t^B$$
$$t = \quad (1.5896) \quad (-1.3958) \quad (-2.1926) \quad (3.3141) \qquad (11.8.13)$$
$$R^2 = 0.9932 \qquad d = 1.7115$$

حيث $\widehat{PPCX}^A$ الموجود في الجانب الأيمن من (12.8.13) تم الحصول عليها من النموذج A، (9.8.13). ولكن في هذا الانحدار، معامل $\widehat{PPCX}_t^A$ الموجود على الجانب

الأيمن معنوي إحصائيًا (عن مستوى 0.0425 من الاختبار ذي الطرفين). وفقًا لخطوات اختبار J، يجب أن نرفض النموذج A في صالح النموذج B.

والآن افترض أننا جعلنا النموذج B هو النموذج المرجعي، والنموذج A موجود في الفرض البديل، وسنتبع نفس الخطوات السابقة، وحصلنا على النتائج التالية:

$$\widehat{PPCE}_t = -6{,}549.8659 + 5.1176PDPI_t + 0.6302PPCE_{t-1} - 4.6776\widehat{PPCE}_t^A$$
$$t = \quad (-2.4976) \quad (2.5424) \quad (3.4141) \quad (-2.1926) \qquad (12.8.13)$$
$$R^2 = 0.9920 \qquad d = 1.7115$$

حيث $\widehat{PPCX}^A$ الموجودة على الطرف الأيمن من (12.8.13) تم الحصول عليها من النموذج A، (9.8.13). ولكن في هذا الانحدار، معامل $\widehat{PPCX}_t^A$ الموجود على الجانب الأيمن معنوي إحصائيًا (عند مستوى 0.0425 من اختبار ذي طرفين). هذه النتائج تعني أنه يجب رفض النموذج B في صالح النموذج A!

كل ذلك يعني أننا غير قادرين على رفض أو قبول أي من النموذجين، وبالتالي لا تستطيع تفسير نفقات الاستهلاك الشخصية في الولايات المتحدة الأمريكية خلال الفترة 1970 - 1991.

بالطبع لقد درسنا نموذجين اثنين فقط. في الواقع قد يوجد أكثر من نموذجين. أسلوب اختبار J يمكن توسيع نطاقه ليشتمل على مقارنة عدد من نماذج الانحدار، إلا أن أسلوب التحليل يصبح معقدًا بسهولة.

هذا المثال يوضح بشكل كبير أسباب افتراض CLRM أن نموذج الانحدار المستخدم في التحليل هو النموذج الصحيح، ومن الواضح أنه عند تكوين النموذج يجب دراسة الظاهرة محل التحليل بشكل دقيق وشامل.

اختبارات أخرى لاختيار النموذج: Other Tests of Model Selection

اختبار نموذج J السابق مناقشته، يعتبر واحدًا من مجموعة من الاختبارات المستخدمة لاختيار النموذج. هناك **اختبار Cox**، **اختبار JA**، **اختبار P والاختبار الشامل لـ Mizon-Richard**، والعديد من الاختبارات الأخرى. بالطبع لن نستطيع مناقشة كل هذه الاختبارات في إطار الكتاب الحالي، وعلى القارئ أن يعود إلى قائمة المراجع المشار إليها في الهوامش المختلفة.[37]

(37) انظر أيضًا في Badi H. Baltagei, Econometrics, Springer, New York, 1998, pp. 209-222.

9.13 معيار اختيار النموذج: MODEL SELECTION CRITERIA

في هذه الفقرة سنناقش المعايير المختلفة التي يمكن على أساسها الاختيار ما بين النماذج المتنافسة، أو مقارنة بين بعض النماذج لأغراض التنبؤ. هنا نرفق بين التنبؤ داخل العينة والتنبؤ خارج العينة. التنبؤ داخل العينة يعبر عن كيفية توفيق أو مناسبة النموذج المختار للبيانات المعطاة في العينة. أما التنبؤ خارج العينة فيعني تحديد كيفية أن يقوم النموذج المختار بالتنبؤ بقيم مستقبلية للمتغير المنحدر عليه بناء على قيم معطاة للمتغيرات المنحدرة.

العديد من المعايير تُستخدم لهذا الفرض. في الواقع لقد ناقشنا بعض هذه المعايير: (1) R^2، (2) R^2 المعدلة ($\bar{R}^2 =$)، (3) معيار Akaike Information Criterion (ATC)، (4) معيار Schwarz Information Criterion (SIC) ، (5) معيار Mallow's C_p و (6) كاي التربيعي المتنبأ به.

كل هذه المعايير تهدف إلى تصغير حجم مجموع مربعات البواقي (RRS) (أو زيادة قيمة R^2). عموماً بخلاف المعيار (1) فإن المعايير (2)، (3)، (4)و (5) تكون هناك عواقب خطيرة لإضافة المزيد أو عدد كبير من المتغيرات المنحدرة. وبالتالي تكون هناك موازنة بين جودة التوفيق للنموذج ودرجة تعقيده (والتي يحكم عليها بناء على عدد المتغيرات المنحدرة الموجودة فيه).

معيار R^2 : The R^2 Criterion

نعرف أن أحد مقاييس جودة التوفيق لنموذج الانحدار هو R^2 ويعرف كالتالي:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}} \tag{1.9.13}$$

وبالتالي R^2 يجب أن تقع بين 0 و 1. كلما اقتربت من 1 كلما كان لدينا توفيق أفضل. ولكن هناك العديد من المشاكل المتعلقة بـ R^2. **أولاً**: أنه يعيش جودة التوفيق داخل العينة، بمعنى قرب القيم المقدرة للـY من قيمها الفعلية الموجودة في العينة المعطاة. وذلك لا يضمن بأي حال من الأحوال القدرة على التنبؤ من خارج العينة. **ثانياً**: في مقارنة اثنين أو أكثر من قيم R^2، يجب أن يكون المتغير التابع أو المنحدر عليه واحداً. **ثالثاً**: والذي يعتبر أكثر أهمية، R^2 لا تقل عند إضافة المزيد من المتغيرات إلى النموذج. وبالتالي من الممكن دائماً "زيادة R^2" بمزيد من المتغيرات المنحدرة المضافة

للنموذج. ولكن بالطبع إضافة مزيد من المتغيرات للنموذج رغم أنها تزيد من قيمة R^2 إلا أنها تزيد أيضًا تباين الخطأ المتوقع.

R^2 المعدلة Adjusted R^2

لتفادي أو تصحيح مسألة زيادة أعداد المتغيرات المنحدرة لزيادة قيمة R^2. قام Henry Theil بعمل R^2 المعدلة، ونرمز لها بالرمز $\bar{R}^2$، والتي سبق وناقشناها في الفصل (7).

تذكر أن:

$$\bar{R}^2 = 1 - \frac{\text{RSS}/(n-k)}{\text{TSS}/(n-1)} = 1 - (1 - R^2)\frac{n-1}{n-k} \qquad (2.9.13)$$

وكما ترى من هذه المعادلة، $\bar{R}^2 \leq R^2$، وبالتالي يتضح تعامل R^2 المعدلة مع إضافة متغيرات جديدة للنموذج. وكما لاحظنا في الفصل (8)، وبخلاف R^2، R^2 المعدلة ستزيد فقط إذا كانت القيمة المطلقة لـ t الخاصة بالمتغير المضاف أكبر من 1. إذا كنت تقارن بين نماذج، يعتبر استخدام $\bar{R}^2$ أفضل من R^2. ولكن مرة أخرى، ضع في الاعتبار أن المتغير المنحدر عليه، يجب أن يكون نفس المتغير بالنسبة للنماذج المتقارن بينها.

معيار المعلومات لـ Akaike (AIC): Akaike Information Criterion (AIC)

فكرة علاج مشكلة إضافة متغيرات منحدرة جديدة للنماذج تم تناولها بشكل أكبر في معيار AIC والمعروف كالتالي:

$$\text{AIC} = e^{2k/n}\frac{\sum \hat{u}_i^2}{n} = e^{2k/n}\frac{\text{RSS}}{\text{n}} \qquad (3.9.13)$$

حيث k تمثل عدد المتغيرات المنحدرة (مشتملة على الجزء الثابت المقلوع من المحور (الصادي) و n تمثل عدد المشاهدات. للتسهيل الرياضي (3.9.13) يمكن كتابتها كالتالي:

$$\ln \text{AIC} = \left(\frac{2k}{n}\right) + \ln\left(\frac{\text{RSS}}{n}\right) \qquad (4.9.13)$$

حيث ln AIC = اللوغاريتم الطبيعي لـ AIC و $2k/n$ = معامل الجزاء. بعض الكتب والحزم الإلكترونية تعرف AIC في صورة تحويلة اللوغاريتم فقط، وبالتالي لا يوجد ln قبل ATC. كما نرى من هذه المعادلة، AIC يفرض جزءًا كبيرًا أكثر من R^2 على زيادة أعداد المتغيرات المنحدرة للمقارنة بين نموذجين أو أكثر. النموذج الذي له قيمة أقل في ATC يعتبر النموذج المفضل. إحدى مميزات استخدام AIC أنه مقيد

ليس فقط في التنبؤ من داخل العينة ولكن خارجها أيضًا، وأيضًا يمكن استخدام في حالة النماذج المتداخلة وغير المتداخلة. ويستخدم أيضًا في تحديد عدد الفترات الزمنية السابقة التي يجب استخدامها في نماذج AR (p).

معيار المعلومات لـ Schwarz (SIC): Schwarz Information Criterion (SIC)

بنفس فكرة AIC، معيار SIC يعرف كالتالي:

$$\text{SIC} = n^{k/n}\frac{\sum \hat{u}^2}{n} = n^{k/n}\frac{\text{RSS}}{n} \tag{5.9.13}$$

أو في الشكل اللوغاريتمي التالي:

$$\ln \text{SIC} = \frac{k}{n}\ln n + \ln\left(\frac{\text{RSS}}{\text{n}}\right) \tag{6.9.13}$$

حيث $[(k/n) \ln n]$ هو معامل الجزاء. SIC لديها معامل جزاء أشد صعوبة وقسوة من AIC ونرى ذلك واضحًا بمقارنة (6.9.13) مع (4.9.13). كما الحال في الـ AIC، فإنه كلما قلت قيمة SIC كلما كان النموذج أفضل. ومرة أخرى، كما الحال في AIC، فإن SIC يمكن أن تستخدم في المقارنة بين أداء النموذج التنبؤى داخل العينة من جهة، وخارجها من جهة أخرى.

معيار Mallows's C_p:

افترض أن لدينا نموذجًا من k متغيرًا مستقلاً مضافًا إليها الجزء الثابت المقطوع من المحور الصادي. ودع $\hat{\sigma}^2$ كعادتها تمثل تقدير لـ σ^2 الحقيقية. ولكن افترض أن لدينا فقط p تغير مستقل حيث $(p \leq k)$ واحصل على RSS من الانحدار باستخدام هذه المتغيرات (p). اجعل RSSp ترمز إلى مجموع مربعات البواقي باستخدام p متغير. وقد قام C.P.Mallows بعمل المعيار التالي لاختيار النموذج والمعروف باسم معيار C_p:

$$C_p = \frac{\text{RSS}_p}{\hat{\sigma}^2} - (n - 2p) \tag{7.9.13}$$

حيث n تمثل عدد المشاهدات.

نعرف أن $E(\hat{\sigma}^2)$ مقدر غير متحيز للمعلمة σ^2 الحقيقية. والآن إذا كان النموذج الذي يحتوي على p متغير نموذج سليم، ولا يعاني من عدم التوفيق، يمكن إثبات أن $E(\text{RSS}_p) = (n-p)\sigma^2$ [38]. وبالتالي يمكن القول بتحقق التالي تقريبًا:

---

(38) Norman D. Draper and Harry Smith, Applied regression Analysis, 3d ed., John Wiley & Sons, New York, 1998 p. 332. انظر أمثلة تطبيقية خاصة بـ C_p.

$$E(C_p) \approx \frac{(n-p)\sigma^2}{\sigma^2} - (n-2p) \approx p \qquad (8.9.13)$$

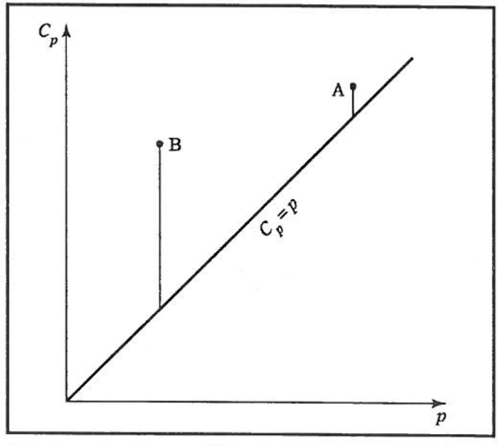
وبالتالي عند اختيار نموذج بناء على معيار C_p، سنختار النموذج الذي له أقل قيمة لـ C_p، تقريبًا المساوية للـ p. بمعنى آخر سنختار نموذجًا يحتوي على p متغير $(p < k)$، والذي يحقق جودة توفيق جيدة للبيانات.

في التطبيق العلمي عادة ما يتم رسم C_p المحسوب من (7.9.13) مع p. والنموذج السليم ستكون فيه نقطة قريبة من الخط الذي تمثله $C_p = p$ ويمكننا رؤية ذلك في الشكل (3.13). وكما يتضح من الشكل، فإن النموذج A يكون مفضلاً عن النموذج B، حيث إن قريبًا من الخط $C_p = p$ عن النموذج B.

تحذير خاص بمعيار اختيار النموذج: A Word of Caution about Model Selection Criteria

سبق وناقشنا العديد من المعايير الخاصة باختيار النموذج، ولكن يجب مراعاة التعامل مع هذه المعايير في ضوء اختبارات التوصيف المختلفة، التي سبق مناقشتها في هذا الفصل. بعض هذه المعايير يعتبر وصفيًا إلى حد كبير، ولا توجد لديه الخصائص النظرية القوية. والبعض الآخر متصل بتنقيب البيانات. وجميع المعايير السابق ذكرها لا يوجد منها ما يستخدم بكثرة من الباحثين، مما يجعل القارئ على دراية مسبقة عنهم. ولا يوجد أحد منهم متفوق على الآخر.[39]

معظم الحزم الإلكترونية الحديثة تشتمل على R^2، R^2 المعدلة، AIC و SIC. ولكن Mallows's C_p ليست متاحة بشكل مباشر وإن كان من السهل تعريفها.



شكل (3.13) الرسم البياني الخاص بـ Mallows's C_p

(39) لقراءة أكثر عمقًا خاصة بهذه النقطة، انظر Francis X. Diebold, Elements of Forecasting, 2d ed., South Western Publishing, 2001, pp. 83-89. وسترى أن Diebold ينصح باستخدام معيار STC.

كاي التربيعي (χ^2) التنبؤي Forecast Chi-Square (χ^2)

افتـرض أن لدينا نموذج انحـدار به n مشـاهدة، ودعنا نفـترض أننا نريد التنبؤ بالقيـمة المتوقـعة للمتـغير المنحـدر عليه بعد إضـافة t مشاهدة - كما سبق ورأينا - فإن فكرة الاحـتفـاظ ببـعض بيـانات العينة لـنرى قـدرة النمـوذج المقـدر على التنبـؤ بهـذه المشاهدات غير الموجودة في القيمة تعتبر فكرة صائبة.

الآن دعنا نعرف اختبار $\chi 2$ التنبؤي كالتالي:

$$\text{Forecast, } \chi^2 = \frac{\sum_{n+1}^{n+t} \hat{u}_i^2}{\hat{\sigma}^2} \tag{9.9.13}$$

حيـث $\hat{u}_i$ هو الخطأ المتوقع في الفتـرة i ($n + 1, n + 2, ..., n + t =$) باسـتخـدام المعلمات التي حـصلنا عليها من نموذج الانحدار المقـدر، وقيم المتغيـرات المنحدرة في الفتـرة التـاليـة للعينة. $\hat{\sigma}^2$ هي القيـمـة المعتـادة لمقـدر الـ OLS الخـاص بـ σ^2 بناء على الانحدار المقدر.

إذا افترضنا أن قيم المعالم لم تتغير ما بين العينة والفترة التالية لها، يمكن إثبات أن الإحصاء المعطى في (9.9.13) يتبع توزيع كاي - التربيعي بدرجات حرية t، حيث إن t هي عدد الفترات التي يتم التنبـؤ فيها. وكما لاحظ Charemza و Deadman فإن اختبار χ^2 التنبؤي له قـوة إحصائية ضعـيفة. بمعنى أن احتـمال رفض الاختبار لفرض خاطئ صغير وبالتالي يجب استخدام الاختبار كضوء أولي وليس كاختبار مؤكد.[40]

10.13 موضوعات إضافية في نمذجة الاقتصاد القياسي ADDITIONAL TOPICS IN ECONOMETRIC MODELING

كمـا لاحظنا في مقـدمة هذا الفـصل، موضـوع نمذجة الاقـتصـاد القيـاسي والاختـبارات المتـعلقـة بتوصيـف النموذج عـديدة ومتنوعة، وتحتـاج إلى كتب متخصـصة في هذه النقطة فقط. في الفـقرة السـابقة حـاولنا أن نتلمس بعض النقاط المهمـة الخاصة بهـذا المجال. في الفقـرة الحاليـة دعنا نستعرض بعض الخـواص التي قد

(40) Wojciech W. Charemza and Derek F. Deadman, New Directions in Econometric Practice: A General to Specific Modelling, Cointegration and Vector Autoregression, 2d ed., Edward Elgar Publishers, 1997, p. 30. See alos pp. 250-252 for their views on various model selection criteria.

يجدها الباحثون مهمة عند التطبيق، وبالأخص سنستعرض التالي: (1) **القيم الشاذة (الفاعلية والتأثير)**، (2) **المربعات الصغرى التكرارية**، (3) **اختبار فشل التنبأ لـ Chow**. مع وضع في الاعتبار أن مناقشتنا لهذه الموضوعات ستكون مناقشات مختصرة.

القيم الشاذة (الفاعلية والتأثير)[41]؛ Outliers, Leverage, and Influence

تذكر أنه عند تصغير مجموع مربعات البواقي (RSS) فإن OLS تعطي أوزانًا متساوية لكل المشاهدات الموجودة في العينة. ولكن ليست كل المشاهدات لها تأثير متساو على نتائج الانحدار، حيث توجد ثلاثة أنواع تعتبر نقاط بيانات خاصة، وتسمى نقاط القيم الشاذة (قيم الفاعلية وقيم التأثير). ومن المهم أن نتعرف على طبيعة هذه النقاط والكيفية التي تؤثر بها على تحليل الانحدار.

في موضوع الانحدار، القيم الشاذة يمكن تعريفها على أنها المشاهدة التي لها "بواق كبيرة". تذكر أن $\hat{u}_i = Y_i - \hat{Y}_i$ أي أن البواقي تمثل الفروق (الموجبة والسالبة) بين القيم الحقيقية للمتغير المنحدر عليه، وقيمته المقدرة من نموذج الانحدار. عندما نجد بواقي كبيرة مقارنةً مع البواقي الأخرى عادة فإنها تلفت نظرنا بسرعة، خصوصًا لبعدها عن خط الانحدار المقدر. لاحظ أنه في مجموعة البيانات الواحدة قد نجد أكثر من قيمة شاذة واحدة. ولدينا مثال متعلق بهذه النقطة في تمرين (2.2.11)، حيث يطلب من القارئ أن يقوم بعمل انحدار لنسبة التغير في أسعار الأسهم (Y) على نسبة التغير في أسعار المستهلك (X) لعينة من 20 دولة. ونجد أن مفردة واحدة، المتعلقة بدولة تشيلي، تعتبر قيمة شاذة.

نقطة البيانات التي لها فاعلية (عالية) تطلق على النقطة التي لا تتناسب مسافة بعدها عن مجموعة قيم المتغير أو المتغيرات المنحدرة. لماذا تعتبر النقاط الفعالة ذات أهمية؟ تعتبر مثل هذه النقاط مهمة، لأنها قادرة على أن تجذب خط الانحدار ناحيتها، وبالتالي تدمر ميل خط الانحدار. وإذا حدث ذلك بالفعل فإننا نطلق على هذه النقطة الفعالة نقطة التأثير. حذف مثل هذه النقاط من العينة يكون له تأثير كبير على خط الانحدار. وبالعودة إلى تمرين (22.11)، سنجد أنه إذا قمت بعمل انحدار لـ Y على X مستخدمًا في التحليل مشاهدة دولة تشيلي، فستحصل على معامل ميل

(41) المناقشة التالية خاصة بـ Chandan Mukherjee. Howard White, and Marc Wyuts, Econometrics and Data Analysis for Developing Countries, Routledge, New York, 1998, pp 137-148.

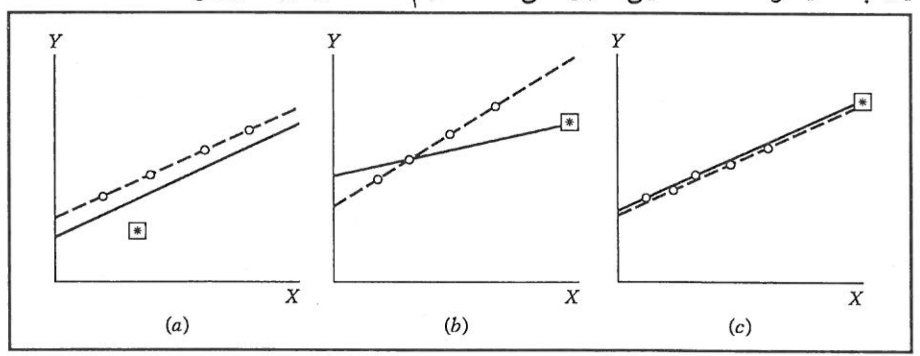
موجب، وله معنوية إحصائية عالية. ولكن إذا حذفت مشاهدة تشيلي، فإن الميل يساوي صفراً، وبالتالي المشاهدة التشيلية تعتبر نقطة فاعلية وتأثير أيضاً.

لتوضيح طبيعة النقاط الشاذة (الفاعلية والتأثير) دعنا نستعرض شكل (4.13) والذي يشرح نفسه بنفسه.[42]

كيف يمكننا التعامل مع مثل هذه النقاط؟ هل نقوم بحذفها ونركز اهتمامنا فقط على باقي نقاط البيانات؟ دعنا نرى إجابة ذلك من خلال Draper و Smith:

الرفض الأتوماتيكي للقيم الشاذة ليس دائماً القرار الحكيم. حيث إن القيم الشاذة أحياناً تعطينا معلومات لا نستطيع الحصول عليها من نقاط البيانات الأخرى، وذلك نتيجة نشأتها في توليفة خاصة قد يكون لها دور وتأثير، وتحتاج المزيد من التحليل وليس الرفض.

كقاعدة عامة، القيم الشاذة يجب رفضها مباشرةً إذا تتبعنا أخطاء خاصة بتسجيل البيانات، أو تحديد المواصفات (في التجارب العملية). بخلاف ذلك، يجب أن يكون هناك تحليل دقيق لمثل هذه القيم الشاذة وطبيعة وجودها.[43]



شكل (4.13): في كل شكل جزيئي، الخط المتصل يمثل خط OLS لكل البيانات، والخط المتقطع يمثل خط OLS بالقيم الشاذة المحذوفة والتي نرمز لها بـ[*]. في (a) القيمة الشاذة قريبة من القيمة المتوسطة لـ X ولها فاعلية قليلة وتأثير قليل على معاملات الانحدار. في (b) القيمة الشاذة بعيدة عن القيمة المتوسطة لـ X ولها فاعلية عالية وتأثير كبير على معاملات الانحدار. في (c)، القيمة الشاذة لها فاعلية عالية ولكن لها تأثير قليل على معاملات الانحدار، حيث إنها على نفس الخط التي تقع عليه باقي المشاهدات.

المصدر: Adapted from John fox, op. cit, p. 268

(42) Adapted from John Fox, Applied Regression Analysis, Linear Models, and Related Methods, Sage Publications, California, 1997, p. 268.

(43) Norman R. Drper and Harry Smith, op. cit., p. 76

ما هي الاختبارات التي يمكن استخدامها لاكتشاف القيم الشاذة وقيم الفاعلية؟ هناك العديد من الاختبارات التي سبق وتناولتها الأدبيات، ولكننا لن نستطيع استعراضها جميعًا هنا، حيث يعتبر ذلك خارج نطاق الكتاب الحالي.(44)

بعض الحزم الإلكترونية مثل Shazam و Microfit لديهما أكواد خاص جاهزة لاكتشاف القيم الشاذة (قيم الفاعلية والتأثير).

المربعات الصغرى التكرارية Recursive Least Squares

في الفصل 8، ناقشنا الاستقرار الهيكلي لنموذج الانحدار الخاص ببيانات سلاسل زمنية. وأوضحنا أن اختبار Chow يمكن استخدامه في هذا الإطار.

وبالأخص دعنا نتذكر ما سبق وذكرناه في هذا الفصل، و الخاص بدالة الادخار البسيطة (الادخار كدالة في الدخل) للولايات المتحدة الأمريكية في الفترة 1970 وحتى 1995. وقد رأينا أن علاقة الادخار بالدخل تغيرت تقريبًا منذ 1982. وبمعرفتنا بنقطة التغير الهيكلي، استطعنا التأكد من نتيجة اختبار Chow.

ولكن ماذا سيحدث إذا لم يكن لدينا معلومات عن نقطة أو (نقاط) التغير الهيكلي؟ هنا يمكن استخدام المربعات الصغرى التكرارية (RELS). الفكرة الأساسية وراء طريقة المربعات الصغرى التكرارية يمكن شرحها من خلال انحدار الادخار - الدخل كالتالي:

$$Y_t = \beta_1 + \beta_2 X_2 + u_t$$

حيث Y = الادخار، و X = الدخل والعينة في الفترة 1970 - 1995. [انظر البيانات في جدول (8.9)].

افترض أننا في البداية استخدمنا بيانات 1970 - 1974 وقدرنا بها دالة الادخار، وحصلنا على مقدرات β_1 و β_2. تم استخدامها بيانات 1970 - 1975 وقدرنا مرة أخرى دالة الادخار، وحصلنا على مقدري المعلمتين السابقتين. ثم استخدمنا بيانات

(44) بعض المصادر المهمة: Alvin C. Rencher, Linear Models in Statistics, John Wiley & Sons, New York, 2000, pp. 219-224; A. C. Atkison, Plots, Transformations and Regression: An Introduction to Graphical Methods of Diagnostic Regression Analysis, Oxford University Press, New York, 1985, Chap. 3; Ashis Sen and Muni Srivastava, Regression Analysis: Theory, Methods, and Applications, Springer-Verlag, New York, 1990, Chap. 8; and John Fox, op. cit., Chap. 11.

1970 - 1976 وأعدنا تقدير نموذج الادخار. بهذا الأسلوب في كل مرة نقوم بإضافة نقطة بيانات جديدة لانحدار Y على X حتى نكون قد استخدمنا كل بيانات العينة المتاحة. وكما تعلم فإن كل انحدار سيعطي مجموعة من التقديرات الجديدة لكل من b1 و b2. إذا قمت برسم قيم هذه المعلمات مع كل مرة تقدير، سترى كيف تتغير هذه القيم المقدرة. فإذا كان النموذج محل الدراسة له استقرار هيكلي، فإن تغيرات القيم المقدرة لهذين المعلمتين سيكون صغيراً وعشوائيًا.

أما إذا كانت القيم المقدرة للمعلمات تتغير بشكل معنوي، سيكون ذلك دليلاً على وجود تغيير هيكلي. RELS بهذا الشكل تعتبر وسيلة فعالة للتعامل مع بيانات السلاسل الزمنية، والتي تمتاز بطبيعة الترتيب الزمني. وتعتبر RELS وسيلة فعالة أيضًا في البيانات المقطعية، حيث تكون البيانات مرتبة بشكل مرتبط بـ " حجم " أو " مقاس " المتغير، مثل حجم التوظيف أو الأصول الخاصة بمشروع ما. في تمرين (30.13) يطلب من القارئ أن يطبق RELS على بيانات الادخار المعطاة في جدول (9.8).

الحزم الإلكترونية مثل Schagam، Eviews و Microfit لدينا أكواد خاصة بتقديرات المربعات الصغرى التكرارية. RELS تحسب أيضًا البواقي التكرارية والتي تعتبر أساسًا للعديد من اختبارات التشخيص المختلفة.(45)

اختبار فشل التنبأ لـ Chow : Chow's Prediction Failure Test

سبق وأن ناقشنا اختبار Chow الخاص بالاستقرار الهيكلي في الفصل (8).

أوضح Chow إمكانية تعديل اختباره ليستخدم في اختبار قوة التنبأ الخاصة بأي نموذج انحدار. مرة أخرى، دعنا نعود إلى مثال الانحدار الخاص بالادخار والدخل في الولايات المتحدة الأمريكية خلال الفترة من 1970 - 1995.

افترض أننا قدرنا انحدار الدخل - الادخار خلال الفترة من 1970 - 1981، وحصلنا على $\hat{\beta}_{1,70-81}$ و $\hat{\beta}_{2,70-81}$ واللذين يعتبران مقدرات للجزء الثالث المقطوع من المحور الصادي والميل بناء على بيانات 1970 - 1981. والآن باستخدام القيم الحقيقية للدخل في الفترة 1970 - 1981 دعنا نتنبأ بقيم الادخار لكل من السنوات من 1982 إلى 1995. المنطق هنا هو إذا افترضنا عدم حدوث تغير هيكلي حقيقي في قيم

(45) لمزيد من التفاصيل، انظر Jack Johnston and John DiNardo, Econometric Methods, 4th ed., McGraw-Hill, New York, 1997, pp. 117-121.

المعالم، فإن قيم الادخار المقدرة للفترة من 1982 - 1995 بناء على المعالم المقدرة من الفترة السابقة يجب ألا تختلف كثيراً عن القيم الحقيقية لها في الفترة التالية. بالطبع إذا حدث فرق كبير بين القيم الحقيقية والمقدرة للادخار في الفترة التالية، فإن ذلك يعتبر مؤشراً لعدم استخدام علاقة الدخل - الادخار لكل فترة البيانات المتاحة.

وسواء كان الفرق بين الادخار الحقيقي والمقدر كبيراً أو صغيراً، فإنه يمكن اختباره باستخدام اختبار F كالتالي:

$$F = \frac{\left(\sum \hat{u}_t^{*2} - \sum \hat{u}_t^2\right)/n_2}{\left(\sum \hat{u}_t^2\right)/(n_1 - k)} \qquad (1.10.13)$$

حيث n_1 = عدد المشاهدات في الفترة الأولى (1970 - 1981) والتي تعتبر فترة الانحدار الأساسية أو المبدئية، n_2 = عدد مشاهدات الفترة الثانية أو التنبؤية، RSS=$\sum \hat{u}_t^{*2}$ في حالة استخدام المعادلة لتقدير كل المشاهدات $(n_1 + n_2)$، و RSS = $\sum \hat{u}_t^2$ في حالة تقدير الـ n_1 مشاهدة الأولى و k عدد من المعالم المقدرة (معلمتين في المثال الحالي). إذا كانت الأخطاء مستقلة تتبع التوزيع الطبيعي، فإن إحصاء F المعطى في (1.10.13) يتبع توزيع F بدرجات حرية n_2 و n_1 على الترتيب. في تمرين (31.13) يُطلب من القارئ تطبيق اختبار فشل التنبؤ لـ Chow لنرى ما إذا كانت علاقة الادخار - الدخل قد تغيرت بالفعل أم لا. لاحظ أننا سبق واستعرضنا التماثل بين هذا الاختبار واختبار χ^2 التربيعي التنبؤي من قبل.

11.13 مثال استنتاجي A CONCLUDING EXAMPLE

مثال: نموذج لتحديد الأجر بالساعة: Example: A Model of Hourly Wage Determination

$$\text{Hwage} = \beta_1 + \beta_2\text{Edu}_i + \beta_3\text{Gender}_i + \beta_4\text{Hispanic}_i + \beta_5\text{Lfexp}_i + \beta_6\text{Mstatus}_i + \beta_7\text{Race}_i + \beta_8\text{Region}_i + \beta_9\text{Union}_i + u_i \qquad (1.11.13)$$

حيث Hwage = الأجر بالساعة ($).

Edu = التعليم بالسنوات.

Gender = 1 انثى، 0 غير ذلك.

Hispanic = 1 Hispanic، 0 غير ذلك.

Race = 1 للعرق غير الأبيض وليس Hispanic، 0 غير ذلك.

Lfexp = خبرة متوقعة في سوق العمل بالسنوات.

Mstatus = الحالة الاجتماعية، 1 متزوج، 0 غير ذلك.

Region = محل الإقامة، 1 للجنوب، 0 غير ذلك.

Union = حالة العمل، 1 إذا كان تبع اتحاد العمل، 0 غير ذلك.

أصل دالة الأجر (1.11.13) يرجع إلى Jacob Mincer.[46] كما ترى فإن دالة الأجر تشتمل على متغيرات كمية، وأخرى نوعية أو وهمية. مبدئيًا، فإن كل هذه المتغيرات تعتبر متغيرات منطقية. ولاحظ أن متغير العرق له ثلاث فئات: إسباني، إسباني غير أبيض، إسباني أبيض.

البيانات تتكون من 528 مفردة تم تجميع البيانات في 1985 كجزء من مسح السكان الحالي (CPS) والذي يقوم به مكتب التعداد الأمريكي كل فترة زمنية معينة. هذه البيانات قام بتجميعها Berndt وتم تكييفها للدراسة بواسطة Goldberg. سبق وأن أشرنا إلى هذا المصدر من قبل في الفصل (2). ودعنا نضع في الاعتبار دائمًا أن هذه البيانات تعتبر بيانات مقطعية.

مبدئيًا، يتوقع أن تكون هناك علاقة طردية بين الأجر بالساعة والتعليم وسنوات الخبرة، الحالة الاجتماعية وحالة العمل من جهة، وتكون هناك علاقة عكسية بين الأجر بالساعة والعرق والنوع ومحل الاقامة. لاحظ أن الفئة المرجعية هي غير الاسباني الأبيض. يمكنك الرجوع إلى أي كتاب خاص باقتصاديات العمل للتعرف على العوامل المختلفة المحددة للأجر بالساعة.[47]

طلبت من تلاميذي بناء على البيانات المتاحة، أن يقوموا بتقدير النموذج (1.11.13). نتائج الانحدار معطاة في جدول (4.13).

وكما ترى، كل المتغيرات في (1.11.13) لها الإشارات المتوقعة، حتى وإن لم تكن جميع المتغيرات لها معنوية إحصائية منفردة.

قيمة R^2 تقترب من 0.2826 والتي قد تبدو قيمة صغيرة نوعًا ما، ولكن عادة ما نرى مثل هذه القيم في حالة البيانات المقطعية التي تشتمل على عدد كثير من المشاهدات. ولكن لاحظ أن قيمة R^2 معنوية، حيث إن قيمة F المحسوبة تساوي تقريبًا 25.56 وهي قيمة لها معنوية عالية، وقيمة P-value تساوي تقريبًا الصفر: تذكر أن إحصاء F يختبر الفرض القائل بأن معاملات الميل جميعًا تساوي الصفر آنيًا، أي أن كل المتغيرات المفسرة في نفس الوقت ليس لها تأثير على المتغير المنحدر عليه.

---

(46) انظر J. Mincer, School, Experience and Earnings, Columbia University Press, New York, 1974.

(47) انظر، على سبيل المثال، George Borjax, Labor Economics, 2d. ed., McGraw-Hill, New York, 2000.

لاحظ لا توجد معنوية إحصائية منفردة لمتغير العرق الإسباني، الحالة الاجتماعية والعرق.

في حين هناك معنوية إحصائية لمتغير محل الإقامة، بعض تلاميذي قاموا بحذف المتغيرات الثلاثة الأولى، وقاموا بعمل الانحدار مرة أخرى، وحصلوا على النتائج الموجودة في جدول (5.13). والآن نرى أن جميع المتغيرات لها معنوية إحصائية منفردة عند مستوى معنوية 5% (أي أن P-value أقل من 5%). تفسير العديد من المتغيرات يأتي بشكل مباشر. فعلى سبيل المثال:

جدول (4.13) نتائج الانحدار بناء على (1.11.13)

Dependent Variable: HWAGE
Sample: 1 528

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.182714	1.275908	-3.278227	0.0011
EDUCATION	0.937130	0.082625	11.34194	0.0000
GENDER	-2.140661	0.391546	-5.467200	0.0000
HISPANIC	-0.512385	0.911056	-0.562408	0.5741
LFEXP	0.098486	0.017494	5.629597	0.0000
MSTATUS	0.485134	0.418881	1.158167	0.2473
RACE	-0.942389	0.583578	-1.614849	0.1070
REGION	-0.771424	0.430173	-1.793287	0.0735
UNION	1.468088	0.512735	2.863248	0.0044

R-squared	0.282693	Mean dependent var	9.047538
Adjusted R-squared	0.271636	S.D. dependent var	5.144082
S.E. of regression	4.390177	Akaike info criterion	5.813515
Sum squared resid	10003.03	Schwarz criterion	5.886283
Log likelihood	-1525.768	F-statistic	25.56745
Durbin-Watson stat	1.857457	Prob(F-statistic)	0.000000

جدول (4.13)

Dependent Variable: HWAGE
Sample: 1 528

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-4.289796	1.258229	-3.409392	0.0007
EDUCATION	0.953006	0.082184	11.59596	0.0000
GENDER	-2.134171	0.391740	-5.447929	0.0000
LFEXP	0.104037	0.016888	6.160545	0.0000
REGION	-0.840832	0.427621	-1.966303	0.0498
UNION	1.427421	0.509978	2.798988	0.0053

R-squared	0.276707	Mean dependent var	9.047538
Adjusted R-squared	0.269779	S.D. dependent var	5.144082
S.E. of regression	4.395772	Akaike info criterion	5.810462
Sum squared resid	10086.51	Schwarz criterion	5.858974
Log likelihood	-1527.962	F-statistic	39.93978
Durbin-Watson stat	1.858629	Prob(F-statistic)	0.000000

القيمة 0.8408 – والخاصة بالمتغير الوهمي المرتبط بمحل الإقامة يمكن تفسيره كالتالي: مع افتراض ثبات باقي العوامل، فإنه في المتوسط العمال في الجنوب يتقاضون في الساعة الواحدة 84 سنتًا أقل من نظرائهم في المناطق الأخرى، ويمكن تعليل ذلك بانخفاض تكاليف الحياة في الجنوب أو الحقيقة الخاصة بطبيعة الجنوب وطبيعته في الاتحاد.

بالمثل، في المتوسط، فإن الإناث تتقاضى أقل من نظرائهن من الرجال بحوالي $2.13 مع افتراض ثبات باقي العوامل الأخرى. بالطبع هذا المقدر من التحيز والتمييز النوعي لا يمكن تحديده فقط بناء على التحليل الإحصائي.

كالمتوقع الانحدار "القصير" الذي تم فيه حذف متغيرات العرق والحالة الاجتماعية له R^2 معدلة أصغر من الانحدار "الطويل" (أي الانحدار الذي يشتمل على كل المتغيرات). ولكن لاحظ أن إحصاء Akaike و Schwarz كلاً منهما صغير في حالة الانحدار القصير، مقارنةً بالانحدار الطويل، مما يوضح أثر إدخال متغيرات أكثر على النموذج. وبما أن الاحصائين الاثنين قريبان في القيمة لبعضهما البعض، فمن الممكن أن تستخدم أيًا منهما، قيمة إحصاء Durbin-Watson في كل من النموذجين تقترب من 2 مما يشير إلى اخطاء "الارتباط الذاتي" أو أخطاء التوصيف.

بما أن البيانات الخاصة بانحدار (1.11.13) متاحة في أسطوانة البيانات الملحقة بالكتاب، فيمكن للقارئ أن يحاول "تجربة" هذه البيانات.

فقد يكون من المحتمل وجود تفاعل بين المتغيرات الوهمية الخاصة بالنوع، والمستوى التعليمي، والحالة الاجتماعية. ومن المحتمل أيضًا أن تكون هناك علاقة غير خطية بين الأجر بالساعة، والخبرة في سوق العمل، مما يتطلب إدخال متغير التعليم كقيمة مربعة في نموذج الانحدار.

كما ترى، حتى وفقًا لبيانات معينة، هناك العديد من الاحتمالات المختلفة. وهذا يجعل الأمر يبدو وكأنه قريب من تنقيب البيانات، وقد سبق أن ذكرنا أن تنقيب البيانات قد يكون له دور ما في نمذجة الاقتصاد القياسي. بالطبع يجب أن نضع في الاعتبار مستوى المعنوية الحقيقي عند استخدام تنقيب البيانات.

12.13 كلمة خاصة للممارس: A WORD TO THE PRACTITIONER

لقد غطينا في هذا الباب العديد من النقاط المتعلقة بالموضوع محل الدراسة، وليس هناك شك من أن بناء النموذج يعتبر فنًا إلى جانب كونه علمًا قائمًا بذاته. الباحث الممارس يجب أن يكون على دراية بالنظرية العلمية ووراء بحثه بالإضافة إلى أدوات التحليل والتشخيص. ولكن من المفيد أن نضع في الاعتبار ملاحظة Martin Feldstein والخاصة بالتالي: «الممارس للاقتصاد القياسي، مثله مثل العالم النظري، يكتشف لاحقًا من الخبرة والتجربة أن النموذج المفيد ليس بالضرورة النموذج " الصحيح " أو " الواقعي " ولكنه قد يكون النموذج المتاح الذي نحصل منه على المعلومات المرجوة».[48]

أما Peter Kennedy من جامعة Simon Fraser في كندا، فقد وضح التالي: **" الوصايا العشر الخاصة بالاقتصاد القياسي التطبيقي "**[49]:

1 – استخدم الحس العام والنظرية الاقتصادية معًا.

2 – اسأل الأسئلة الصحيحة (أي المتعلقة بالموضوع حتى وإن كانت غير رياضية).

3 – تعرَّف جيدًا على الموضوع محل البحث (لا تهتم فقط بالجانب الإحصائي).

4 – تحقق من البيانات.

5 – لا تعقد الأمور المتعلقة بالبحث. استخدم مبدأ "KISS" **KISS Principle** والذي يتطلب تبسيط كل الأمور العشوائية.

6 – انظر بدقة وببعد نظر ولفترة طويلة على النتائج.

7 – حدد بشكل حاسم تكلفة تنقيب البيانات.

8 – استعد لعمل موازنات مختلفة (لا تعتمد فقط على ما هو مكتوب في الكتب النظرية).

9 – لا تخلط بين الأهمية والمعنوية (لا تخلط بين المعنوية الإحصائية والمعنوية العملية).

10 – اعترف بوجود الحساسية (أي انتقاد المشارك).

قد يرغب القارئ في قراءة ورقة Kennedy كاملة للتعرف أكثر على هذه الوصايا العشر، والتي قد تبدو للوهلة الأولى غير مستساغة، لكنها عند التطبيق العملي صحيحة إلى حد بعيد.

---

(48) Martin S. Feldstein, "Inflation, Tax Rules and Investment: Some Econometric Evidence," Econometrica, vol. 30, 1982, p. 829.

(47) Peter Kennedy, op. cit., pp. 17-18.

13.13 الخلاصة والاستنتاجات: SUMMARY AND CONCLUSIONS

1 – فرض CLRM والقاضي بأن نموذج الاقتصاد القياسي المستخدم في التحليل يعتبر نموذجًا صحيحًا له معنيان: **الأول**: لا توجد معادلة أخطاء توصيف **وثانيًا**: لا توجد أخطاء توصيف في النموذج. في هذا الفصل تم التركيز على أخطاء توصيف النموذج.

2 – معادلة أخطاء التوصيف التي تم استعراضها في هذا الفصل هي كالتالي:

(1) حذف متغيرات مهمة من النموذج. (2) أثر وجود متغيرات زائفة.
(3) استخدام شكل دالي خاطئ. (4) توصيف غير سليم لحد الخطأ u_i.
(5) أخطاء في القياس سواء في المتغيرات المنحدرة أو المتغير المنحدر عليه.

3 – عندما يتم حذف متغير مهم من النموذج، فإن توابع ذلك تكون خطيرة كالتالي: مقدرات OLS للمتغيرات الموجودة في النموذج لا تكون فقط مقدرات متحيزة، ولكن غير متسقة أيضًا. بالإضافة إلى أن التباين والأخطاء المعيارية الخاصة بهذه المعاملات تكون غير صحيحة، مما يؤثر على خطوات اختبارات الفروض الخاصة بها.

4 – أما التوابع المرتبطة بوجود متغيرات غير مهمة في النموذج، فتكون أقل خطورة كالتالي: مقدرات المتغيرات غير المهمة في النموذج تظل مقدرات غير متحيزة ومتسقة أيضًا، ويكون تقدير تباين الخطأ سليمًا. المشكلة الوحيدة في تقدير التباين أنه يكون أكبر من المفروض، مما يعطي دقة أقل لتقديرات المعالم. أي أن فترات الثقة تكون أوسع من المفروض.

5 – لاكتشاف معادلة أخطاء التوصيف، علينا أن تستخدم أحد الاختبارات التالية مثل (1) اختبار البواقي، (2) إحصاء d لـ Durbin-Watson، (3) اختبار RESET لـ Ramsey، (4) اختبار مضروب Lagrange.

6 – نوع خاص من أخطاء التوصيف مرتبط بالخطأ في القياس في قيم المتغير المنحدر عليه والمتغيرات المنحدرة. إذا كان هناك أخطاء في القياس في المتغير المنحدر عليه فقط، فإن مقدرات OLS تكون غير متحيزة ومتسقة أيضًا، ولكنها أقل في الكفاءة. إذا كانت أخطاء القياس موجودة في المتغيرات المنحدرة فإن مقدرات OLS تكون متسقة وغير متسقة.

7 – حتى وإذا كان اكتشاف أخطاء القياس ممكن، إلا أن كيفية التعامل مع ذلك يظل أمرًا غير سهل. استخدام المتغيرات المساعدة عادة ما يكون جذابًا نظريًا لكن غير

سهل علميًا. حيث إنه من المهم جدًا في الجانب التطبيقي، أن يذكر الباحث صراحةً مصدر البيانات، وطريقة تجميعها، وكيفية تعريف المتغيرات محل الدراسة.

معظم البيانات التي تم تجميعها عن طريق مؤسسات ما يكون فيها هوامش متعددة، لتوضيح أخطاء البيانات وعلى الباحث لفت نظر القارئ لمثل هذه الهوامش.

8 – أخطاء التوصيف غير السليم للنموذج، قد تكون خطيرة مثل معادلة أخطاء التوصيف، في الواقع، يتم التفرقة بين النماذج الشبكية والنماذج غير الشبكية. لاختيار النموذج الملائم، فقد استعرضنا النماذج غير الشبكية، اختبار F، واختبار J لـ Davidson-MacKinnon وأوضحنا أوجه القصور في كل اختيار على حدة.

9 – لاختيار نموذج فعلي عند التطبيق العملي، يستخدم الباحثون العديد من المعايير. وقد ناقشنا بعضًا منها مثل Akaike و Schwarz و C_p لـ Mallows وكاي التربيعي التنبؤي. وقد أوضحنا للقارئ أنه لا يوجد معيار واحد من هذه المعايير يمكن اعتباره معياراً مطلقًا ولابد من توخي الحذر عند استخدامها في التحليل.

10 – ناقشنا أيضًا المواضيع التالية: (1) القيم الشاذة (قيم الفاعلية والتأثير)، (2) المربعات الصغرى التكرارية، (3) اختبار فشل التنبؤ لـ Chow. واستعرضنا دور كل منها عند التطبيق العملي.

11 – وخلاصة هذا الفصل، يمكن استنتاجها من الوصايا العشر في الاقتصاد القياسي التطبيقي لـ Peter Kennedy. الفكرة الرئيسة وراء هذه الوصايا هي أن ينظر الباحث إلى ما هو وراء المفاهيم الفنية للاقتصاد القياسي.

تمارين: EXERCISES

أسئلة: Questions

1.13 بالعودة إلى دالة الطلب على الدجاج المقدرة في المعادلة (23.7.8). ومع الوضع في الاعتبار أداء النموذج الجيد الذي تمت مناقشته في الفقرة (1.13). هل يمكنك القول بأن دالة الطلب تعتبر موصفة بشكل "سليم"؟

2.13 افترض أن النموذج الصحيح هو:

$$Y_i = \beta_1 X_i + u_i \qquad (1)$$

ولكننا بدلاً من توفيق هذا النموذج ليمر خلال نقطة الأصل، تم استخدام النموذج التالي الذي يشتمل على جزء ثابت مقطوع من المحور الصادي:

$$Y_i = \alpha_0 + \alpha_1 X_i + v_i \qquad (2)$$

استعرض التوابع الخاصة بمثل هذا النوع من أخطاء التوصيف.

3.13 بالرجوع إلى تمرين (2.13)، افترض أن النموذج (2) هو النموذج الصحيح. استعرض توابع توفيق النموذج الخاطئ (1).

4.13 افترض أن النموذج التالي هو النموذج "الصحيح"

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \qquad (1)$$

ولكن افترضنا أننا أضفنا المتغير X_3 وهو تغيير متغير "غير المهم" في النموذج (غير مهم بمعنى أن المعامل β_3 المرتبط بالمتغير X_3 يساوي الصفر) وقدرنا:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i \qquad (2)$$

(a) هل قيمة R^2 و R^2 المعدلة للنموذج (2) ستكون أكبر من نظيرتها في النموذج (1)؟

(b) هل مقدرات β_1 و β_2 التي حصلنا عليها من (2) غير متحيزة؟

(c) هل وجود المتغير X_3 "غير المهم" له أثر على تباين كل من $\hat{\beta}_1$ و $\hat{\beta}_2$ ؟

5.13 اعتبر دالة الإنتاج التالية لـ Cobb-Douglas واعتبر أنها الدالة "الصحيحة":

$$\ln Y_i = \alpha_0 + \alpha_1 \ln L_{1i} + \alpha_2 \ln L_{2i} + \alpha_3 \ln K_i + u_i$$

حيث Y = الناتج.

L_1 = إنتاج عمالي.

L_2 = إنتاج غير عمالي.

K = رأس المال.

ولكن افترض أن الانحدار الذي تم استخدامه عمليًا هو:

$$\ln Y_i = \beta_0 + \beta_1 \ln L_{1i} + \beta_2 \ln K_i + u_i$$

على افتراض أن لدينا بيانات مقطعية للمتغيرات محل الدراسة.

(a) هل $E(\hat{\beta}_1) = \alpha_1$ و $E(\hat{\beta}_2) = \alpha_3$ ؟

(b) هل إجابة السؤال (a) ستكون صحيحة أيضًا إذا علمنا أن L_2 متغير غير مهم في دالة الإنتاج؟ وضح الخطوات اللازمة لإجابتك.

6.13 بالرجوع إلى المعادلة (4.3.13) و (5.3.13). كما ترى فإن $\hat{\alpha}_2$ على الرغم من أنه متحيز إلا أن تباينه أقل من $\hat{\beta}_2$، والذي يعتبر مقدراً غير متحيز. كيف يمكنك الاختيار بينهما؟

ملاحظة: MSE (متوسط مربعات الأخطاء) للمقدرين الآتيين التالي:

$$\text{MSE}(\hat{\alpha}_2) = \left(\sigma^2 \Big/ \sum x_{2i}^2\right) + \beta_3^2 b_{32}^2$$

تباين العينة + مربع التحيز =

$$\text{MSE}(\hat{\beta}_2) = \sigma^2 \Big/ \sum x_2^2 (1 - r_{23}^2)$$

للمزيد من التفاصيل عن MSE، انظر ملحق A.

7.13 اثبت أن قيمة β المقدرة سواء من (1.5.13) أو (3.5.13) تعتبر مقدراً غير متحيز لقيمة β الصحيحة.

8.13 وفقًا لفرض الدخل الدائم الخاص بـ Friedman يمكننا كتابة التالي:

$$Y_i^* = \alpha + \beta X_i^*$$

حيث Y_i^* = نفقات الاستهلاك الدائمة و X_i^* = الدخل الدائم. افترض أننا بدلاً من ذلك فقد تعاملنا مع التالي:

$$Y_i = Y_i^* + u_i$$

$$X_i = X_i^* + v_i$$

حيث Y_i و X_i هي الكميات التي يمكن قياسها و u_i و v_i هي أخطاء القياس بالترتيب.

باستخدام هذه الكميات، يمكننا كتابة دالة الاستهلاك كالتالي:

$$Y_i = \alpha + \beta(X_i - v_i) + u_i$$

$$= \alpha + \beta X_i + (u_i - \beta v_i) \qquad (2)$$

افترض أن: (1) $E(u_i) = E(v_i) = 0$ ، (2) $\text{var}(u_i) = \sigma_u^2$ و $\text{var}(v_i) = \sigma_v^2$ ،

(3) $\text{cov}(Y_i^*, u_i) = 0$ ، $\text{cov}(X_i^*, v_i) = 0$ ،

و (4) $\text{cov}(u_i X_i^*) = \text{cov}(v_i, Y_i^*) = \text{cov}(u_i, v_i) = 0$

اثبت أنه في العينات كبيرة الحجم، فإن القيمة المقدرة لـ β من (2) يمكن كتابتها كالتالي:

$$\text{plim}(\hat{\beta}) = \frac{\beta}{1 + (\sigma_v^2/\sigma_{X^*}^2)}$$

(a) ما الذي يمكنك قوله بخصوص التحيز في $\hat{\beta}$ ؟

(b) إذا زاد حجم العينة بشكل كاف، هل تقترب القيمة المقدرة لـ β من قيمتها الحقيقية؟

9.13 نموذج السعر لأصول رأس المال. نموذج السعر لأصول رأس المال (CAPM)، والخاص بنظرية الاستثمار الحديثة يفترض العلاقة التالية بين متوسط معدل العائد من الأمن (السهم المشترك) مقاس خلال فترة زمنية معينة، ومدى التباين الأمني، ويُسمى ذلك معامل بيتا (التباين يعتبر مقياسًا للخطر):

$$\bar{R}_i = \alpha_1 + \alpha_2(\beta_i) + u_i$$

حيث $\bar{R}_i$ = متوسط العائد الأمني i.

β_i = معامل بيتا الحقيقية للأمن i.

u_i = حد خطأ عشوائي.

قيمة β_i الحقيقية لا تتم مشاهدتها مباشرةً، ولكن تقاس كالتالي:

$$r_{it} = \alpha_1 + \beta^* r_{m_t} + e_t \quad (2)$$

حيث r_{it} = معدل العائد الأمني i في الفترة t.

r_{m_t} = معدل العائد السوقي للزمن t (هذا المعدل هو معدل خاص ببعض مؤشرات السوق مثل مؤشر S&P)

e_t = حد البواقي

و β^* هي تقدير لمعلمة بيتا "الحقيقية". في الواقع وبدلاً من تقدير (1) فإنه تم تقدير:

$$\bar{R}_i = \alpha_1 + \alpha_2(\beta_i^*) + u_i \quad (3)$$

حيث β_i^* تم الحصول عليها من الانحدار (2). وحيث إن β_i^* مقدرة فإن العلاقة بين β الحقيقية و β^* يمكن كتابتها كالتالي:

$$\beta_i^* = \beta_i + v_i \quad (4)$$

حيث v_i يطلق عليه خطأ في القياس.

(a) ما هو أثر هذا الخطأ في القياس على تقدير α_2 ؟

(b) هل قيمة α_2 المقدرة من (3) تعتبر مقدراً غير متحيز لقيمة α_2 الحقيقية؟ وإذا كانت إجابتك بالرفض، هل تعتبر مقدراً متسقاً لـ α_2؟ وإذا كانت إجابتك بالرفض أيضاً، فما هو العلاج الذي يمكن أن تقترحه لمثل هذه المشكلة؟

10.13 اعتبر النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i \qquad (1)$$

لمعرفة ما إذا كان هذا النموذج موصفاً بشكل غير سليم كنتيجة لحذف المتغير X_3 من النموذج، قررت أن تقوم بعمل انحدار للبواقي التي حصلت عليها من هذا النموذج على (1) المتغير X_3 فقط (لاحظ أن: هناك جزءاً ثابتاً مقطوعاً من المحور الصادي موجود في النموذج). اختبار المضروب Lagrange (LM) يتطلب أن تقوم بعمل الانحدار للبواقي الخاصة بنموذج (1) على كل من X_2 و X_3 وثابت لماذا تعتبر طريقتك في الأغلب طريقة غير ملائمة؟(*)

11.13 اعتبر النموذج التالي:

$$Y_i = \beta_1 + \beta_2 X_i^* + u_i$$

في الواقع تم قياس X_i^* بدلاً من X_i كالتالي:

(a) $X_i = X_i^* + 5$

(b) $X_i = 3X_i^*$

(c) $X_i = (X_i^* + \varepsilon_i)$، حيث i حد خطأ عشوائي له باقي الخواص المعتادة.

ما هو أثر كل من أخطاء القياس السابقة على القيم المقدرة لكل من β_1 و β_2؟

12.13 بالعودة إلى انحدار (1.3.13) و (2.3.13). كمال الحال في (3.3.13) اثبت أن:

$$E(\hat{\alpha}_1) = \beta_1 + \beta_3(\bar{X}_3 - b_{32}\bar{X}_2)$$

حيث $\beta_{3\,2}$ هو معامل الميل في الانحدار الذي تم فيه حذف المتغير X_3 واستخدام المتغير X_2.

13.13 قيم بشكل دقيق وجهة النظر التالية لـ Learner (†)

(*) انظر Maddala, op. cit., p. 477 .

(†) Edward E. Leamer, Specification searches: ad Hoc Inference with Nonexperimental Data, John Wiley & Sons, New York, 1978, p. vi.

اهتمامي بـ metastatistics (أي نظرية الاستدلال المعتمدة فعليًا على البيانات) ينبع من ملاحظتي للاقتصاديين في أوقات العمل. الرأى القائل بأن النظرية الاقتصادية غير مهمة مؤمن به العديد من الممارسين الفعليين للاقتصاد. الفجوة الكبيرة بين النظرية الاقتصادية والتطبيق العملي للاقتصاد متوقع أن تتسبب في ألم كبير لدارسي وباحثي الاقتصاد. في الحقيقة، القليل من التجارب العملية يتماشى مع النظريات المختلفة. مما يجعلنا ننقسم إلى شخصين: شخص يحاول تطبيق النظريات والأساليب الإحصائية، والآخر يحاول أن يحلل البيانات بشكل واقعي وفعلي.

14.13 قيِّم العبارة التالية لـ Henry Theil (*):

وفقًا لطبيعة التحليل الفنية، فإن أفضل طريقة لتفسير فترات الثقة وحدودها المعنوية يجب أن يكون من خلال فترات الثقة واختبارات الفروض المحسوبة بناء على نتائج الانحدار النهائية. فمثلاً 95% فترة ثقة قد تكون في الحقيقة وبناء على النتائج الفعلية لا تمثل أكثر من 80% درجة ثقة ومستوى المعنوية المساوي لـ 1% قد يكون فعليًا 10%.

15.13 وفقًا لطرق الاقتصاد القياسي التي تم تطبيقها في 1950 وبداية 1960، أوضح Blaug التالي: (†)

الكثير منه (أي الأبحاث العملية) يعتبر مثل لعبة التنس، التي تكون فيها شبكة اللعب في مستوى منخفض: فبدلاً من محاولة اختبار التنبؤ ومدى دقته، فإن الاقتصاديين الجدد يحاولون إخضاع الواقع الفعلي لتنبؤاتهم، وبالتالي يستبدلون التقصير، الصعب بالتأكيد السهل.

هل توافق على وجهة النظر السابقة؟ قد تحتاج لقراءة المزيد من تفاصيل وجهة نظر Blaug في كتابه الخاص بذلك.

16.13 وفقًا لـ Blaug، "لا يوجد منطق للإثبات ولكن هناك منطقًا لعدم الإثبات" (‡). ما الذي يعنيه ذلك؟

(*) Henry Theil, Principles of Econometrics, John Wiley & Sons, New York, 1971, pp. 605-606.

(†) M.Blaug, The Methodology of Economics, Or How Economists Explain, Cambridg University Press, New York, 1980, p. 256.

(‡) Ibid., p. 14.

17.13 بالعودة إلى نموذج St. Louis المناقش من قبل، ومع الوضع في الاعتبار المشكلة المتعلقة باختبار F الشبكي، ناقش بدقة النتائج المقدمة في انحدار (4.8.13).

18.13 افترض أن النموذج الصحيح هو:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_2 X_i^2 + \beta_3 X_i^3 + u_i$$

ولكن قدرنا النموذج التالي:

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i$$

إذا لاحظنا Y عند $X = -3, -2, -1, 0, 1, 2, 3$ وقدرنا النموذج "غير السليم" ما هو التحيز الذي سنحصل عليه في نتائج تلك التقديرات؟(*).

19.13 إذا أردنا معرفة ما إذا كان المتغير X_i^2 يتبع النموذج $Y_i = \beta_1 + \beta_2 X_i + u_i$، فإن اختبار Ramsey يمكنه تقدير النموذج الخطي والحصول على قيم Y_i المقدرة من النموذج (أي $\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$) ثم تقدير النموذج $Y_i = \alpha_1 + \alpha_2 X_i + \alpha_3 \hat{Y}_i^2 + v_i$ واختيار معنوية α_3. أثبت أنه إذا كانت $\hat{\alpha}_3$ معنوية في معادلة (RESET) السالبة، فإن ذلك يكون مثل تقدير النموذج التالي مباشرة: $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i$ (ملاحظة: بدل $\hat{Y}_i$ في انحدار RESET).(†)

20.13 وضح مع التعليل ما إذا كانت العبارات التالية صح أم خطأ.(‡)

(a) يمكن اعتبار أحد القيم قيمة مؤثرة لكنها ليست قيمة شاذة.

(b) يمكن اعتبار أحد القيم قيمة شاذة ولكنها ليست قيمة مؤثرة.

(c) يمكن أن تكون هناك قيمة شاذة ومؤثرة في نفس الوقت.

(d) إذا كان النموذج $Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + u_i \hat{\beta}_3$ له معنوية إحصائية، فإنه يجب الاحتفاظ بالجزء الخطي لـ X_i حتى وإن كان $\hat{\beta}_2$ غير معنوي إحصائيًا.

(e) إذا قدرنا النموذج $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ أو $Y_i = \alpha_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ باستخدام OLS، فإن خط الانحدار المقدر سيكون متساويًا في الحالتين حيث تكون $x_{2i} = (X_{2i} - \bar{X}_2)$ و $x_{3i} = (X_{3i} - \bar{X}_3)$.

(*) مختارة من G. A. F. Linear Regression Analysis, John Wiley & Sons, New York, 1977, p. 176.

(†) مختارة من Kerry Peterson, op. cit., pp. 184-185.

(‡) مختارة من Norman R. Draper and Harry Smith, op. cit., pp. 606-607.

مسـائـل : Problems

21.13 استخدم بيانات الطلب على الدجاج المعطاة في تمرين (19.7) وافترض أن دالة الطلب الصحيحة كالتالي :

$$\ln Y_t = \beta_1 + \beta_2 \ln X_{2t} + \beta_3 \ln X_{3t} + \beta_6 \ln X_{6t} + u_t \qquad (1)$$

ولكن الدالة التي تم تقديرها هي كالتالي :

$$\ln Y_t = \alpha_1 + \alpha_2 \ln X_{2t} + \alpha_3 \ln X_{3t} + v_t \qquad (2)$$

حيث Y = استهلاك الدجاج بالنسبة للفرد (Lb).

X_2 = الدخل الحقيقي للفرد.

X_3 = السعر الحقيقي للدجاج.

X_6 = السعر الحقيقي لبدائل الدجاج.

(a) قم بعمل اختبارات RESET و LM لأخطاء التوصيف، وافترض أن دالة الطلب (1) هي الدالة الصحيحة.

(b) افترض أن $\hat{\beta}_6$ الموجود في (1) غير معنوي إحصائيًا. هل يعني ذلك عدم وجود أخطاء توصيف إذا قدرنا النموذج (2) للبيانات المتاحة؟

(c) إذا كان $\hat{\beta}_6$ غير معنوي، هل يعني ذلك ضرورة عدم إدخال المتغير الخاص بسعر بدائل الدجاج كمتغير مضاف لدالة الطلب؟

22.13 بالعودة إلى تمرين 13-21. افترض أن النموذج (2) هو النموذج الصحيح لدالة الطلب :

(a) إذا قدرنا الآن النموذج (1). ما هو نوع خطأ التوصيف الذي قد نقع فيه؟

(b) ما هي التوابع النظرية لهذا النوع من خطأ التوصيف؟ وضح إجابتك وفقًا للبيانات المتاحة .

23.13 افترض أن النموذج الصحيح هو :

$$Y_i^* = \beta_1 + \beta_2 X_i^* + u_i \qquad (1)$$

ولكن نظرًا لوجود أخطاء في القياس، فإنه تم تقدير النموذج التالي :

$$Y_i = \alpha_1 + \alpha_2 X_i + v_i \qquad (2)$$

حيث $Y_i = Y_i^* + \varepsilon_i$ و $X_i = X_i^* + w_i$ ، حيث ε_i و w_i أخطاء قياس.

باستخدام البيانات المعطاة في جدول (2.13). حدد التوابع التالية لتقدير النموذج (2) بدلاً من النموذج الصحيح (1).

24.13 في تمرين 14.6 تم سؤال القارئ عن تقدير المرونة التبادلية بين العمالة ورأس المال باستخدام الـ CES (المرونة الثانية التبادلية) الخاصة بدالة الإنتاج.

ولكن الدالة الموضحة معتمدة على فرض وجود منافسة تامة في سوق العمل، إذا كانت المنافسة غير تامة، فإن الصيغة الصحيحة للنموذج ستكون كالتالي:

$$\log\left(\frac{V}{L}\right) = \log\beta_1 + \beta_2 \log W + \beta_3 \log\left(1 + \frac{1}{E}\right)$$

حيث $\left(\frac{V}{L}\right)$ = القيمة المضافة لكل وحدة عمالة.

L = مدخل العمالة.

W = معدل الأجر الحقيقي.

E = مرونة المعروض من العمالة.

(a) ما نوع خطأ التوصيف الموجود في تقدير CES الأصلي للمرونة التبادلية إذا كان سوق العمل غير تام؟

(b) ما هي العواقب النظرية لهذا الخطأ بالنسبة لـ β_2، أي معامل المرونة التبادلية؟

(c) افترض أن مرونة المعروض من العمالة في هذا المجال الموضح في تمرين (23.6) كان كالتالي: 2، 1.8، 2.5، 2.3، 1.9، 2.1، 1.7، 2.7، 2.2، 2.1، 2.9، 2.8، 3.2، 2.9 و 3.1. باستخدام هذه البيانات مضاف إليها البيانات الموجودة في تمرين (14.6)، قدر النموذج المطلوب؟ وعلق على النتائج في ضوء نظرية أخطاء التوصيف.

25.13 تجربة Monte Carlo: (*) افترض أن لدينا عشرة مشاهدات لها الدخل الدائم التالي: 200$، 220، 240، 260، 280، 300، 320، 340، 360، 380 و 400.

افترض أن الاستهلاك الدائم (Y_i^*) مرتبط بالدخل الدائم X_i^* كالتالي:

$$Y_i^* = 0.8\,X_i^* \qquad (1)$$

كل مفردة من هذه المفردات لها دخل مؤقت يساوي 100 مضروبًا في u_i (ورقم عشوائي) مسحوب من مجتمع يتبع توزيعًا طبيعيًا بمتوسط = 0 وتباين $\sigma^2 = 1$ (أي متغير يتبع توزيعًا معتادًا قياسيًا). افترض أنه لا يوجد جزء في الاستهلاك يعبر عن هذا المقدار المؤقت. وبالتالي الاستهلاك المقاس والاستهلاك الدائم يعتبران نفس الشيء.

(*) مختارة من Christopher Dougherty, Introduction to Econometrics, Oxford University Press, New York, 1992, pp. 253-256.

(a) اسحب 10 أرقام عشوائية من مجتمع يتبع التوزيع الطبيعي بمتوسط يساوي 0 ، وتباين يساوي الوحدة، واحصل على 10 أرقام من الدخل المقاس X_i أي $(= X_i^* + 100u_i)$.

(b) قم بعمل انحدار للاستهلاك الدائم على الدخل المقاس باستخدام البيانات التي حصلت عليها في (a) وقارن نتائجك مع تلك التي حصلت عليها في (1). مبدئيًا يجب أن يكون الجزء الثابت المقطوع من المحور الصادي مساويًا للصفر. لماذا؟. هل حدث ذلك بالفعل؟ علل إجابتك.

(c) كرر (a) 100 مرة، واحصل على 100 انحدار كالموضح في (b) وقارن نتائجك مع تلك التي حصلت عليها من (1). ما هو الاستنتاج العام الذي توصلت إليه؟

26.13 بالرجوع إلى تمرين (26.8) وباستخدام المتغيرات وتعريفتها المعطاة، اعتبر النموذجين التاليين:

نموذج A : $Y_t = \alpha_1 + \alpha_2 X_{3t} + \alpha_3 X_{4t} + \alpha_4 X_{6t} + u_t$

نموذج B : $Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 Y_{5t} + \beta_4 X_{6t} + u_t$

باستخدام اختبار F الشبكي (التداخلي)، كيف يمكنك الاختيار بين النموذجين؟

27.13 بالرجوع لتمرين (26.13). وباستخدام اختبار J، كيف يمكنك الاختيار بين النموذجين؟

28.13 بالعودة إلى تمرين (19.7)، أي من النماذج الخمسة المعطاة أكثر تعبيرًا عن الطلب على الدواجن في الولايات المتحدة الأمريكية؟

(a) ما الفرق بين نموذج (1) ونموذج (2)؟ إذا كان نموذج (2) هو النموذج الصحيح، ولكنك قمت بتقدير النموذج (1)، ما هو نوع الخطأ الذي وقعت فيه؟ ما هو الاختبار الذي يمكنك استخدامه لتحديد معادلة خطأ التوصيف أو خطأ توصيف النموذج؟ وضح الخطوات الحسابية الضرورية لذلك.

(b) من بين النماذج (1) حتى (5). أي النماذج تختار؟ وأي اختبار أو اختبارات ستستخدم؟ ولماذا؟

29.13 بالعودة إلى جدول (9.8) والذي يعطينا بيانات عن الادخار الشخصي (Y) والدخل الشخصي (X) للفترة من 1970 حتى 1995 اعتبر التالي:

النموذج A : $Y_t = \alpha_1 + \alpha_2 X_t + \alpha_3 X_{t-1} + u_t$

النموذج B : $Y_t = \beta_1 + \beta_2 X_t + \beta_3 X_{t-1} + u_t$

كيف يمكنك الاختيار بين النموذجين؟ وضح بشكل دقيق خطوات الاختبار اللازم لذلك، وكل الحسابات اللازمة لذلك الاختبار. افترض أن شخصًا ما اقترح أن متغير سعر الفائدة يعتبر متغيراً ضرورياً لدالة الادخار. كيف يمكنك اختبار ذلك؟ اجمع بيانات عن معدل الفائدة الربع سنوي كمتغير مساعد للفائدة واستخدمه لتعليل إجابتك.

30.13 استخدم بيانات تمرين (29.13). باستخدام المربعات الصغرى التكرارية قدر دالة الادخار خلال 1970 – 1981 و 1971 – 1985 و 1970 – 1990 و 1970 – 1995. علق على استقرار المعاملات المقدرة لدالة الادخار.

31.13 بالرجوع لتمرين (30.13). افترض أنك قدرت دالة الادخار للفترة 1970 – 1981، باستخدام المعلمات المقدرة وبيانات الدخل الشخصي خلال 1982 – 1995، تنبأ بالادخار في فترات لاحقة واستخدم اختبار فشل التنبؤ لـ Chow لتحديد قرار رفض أو قبول الفرض الخاص بعدم حدوث تغيير في دالة الادخار خلال الفترتين.

32.13 حذف متغير من نموذج انحدار يشتمل على K متغير. بالعودة إلى (3.3.13)، والتي توضح التحيز الناتج عن حذف المتغير X_3 من النموذج $Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$، يمكن تعميم ذلك كالتالي: في نموذج يشتمل على K متغير كالتالي: $Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i$، افترض أننا حذفنا المتغير X_k. يمكن إثبات أن تحيز المتغير المحذوف لمعامل الميل الخاص بالمتغير المتبقي عليه X_i هو:

$$E(\hat{\beta}_j) = \beta_j + \beta_k b_{kj} \qquad j = 2, 3, \dots, (k-1)$$

حيث b_{ki} هو معامل الميل (الجزيئي) لـ X_i في الانحدار المساعد الذي تم فيه حذف المتغير X_k من المتغيرات المفسرة الموجودة في النموذج.

بالعودة إلى تمرين (21.13). اوجد تحيز المعلمات الموجودة في المعادلة (1) إذا قمنا بحذف المتغير $\ln X_6$ من النموذج. هل هذا الحذف مؤثر؟ وضح جميع الخطوات الحسابية اللازمة لتوضيح إجابتك.

ملحق A13 APPENDIX

1.A13 إثبات أن $E(b_{1\,2}) = \beta_2 + \beta_3 b_{3\,2}$ [المعادلة (3.3.13)]

THE PROOF THAT $E(b_{1\,2}) = \beta_2 + \beta_3 b_{3\,2}$ [EQUATION (13.3.3)]

انحرافات نموذج الانحدار الخاص بمجتمع يشتمل على ثلاثة متغيرات يمكن كتابتها كالتالي:

$$y_i = \beta_2 x_{2i} + \beta_3 x_{3i} + (u_i - \bar{u}) \tag{1}$$

أولاً بالضرب في x_2 ثم x_3 نحصل على المعادلات الطبيعية التالية:

$$\sum y_i x_{2i} = \beta_2 \sum x_{2i}^2 + \beta_3 \sum x_{2i} x_{3i} + \sum x_{2i}(u_i - \bar{u}) \tag{2}$$

$$\sum y_i x_{3i} = \beta_2 \sum x_{2i} x_{3i} + \beta_3 \sum x_{3i}^2 + \sum x_{3i}(u_i - \bar{u}) \tag{3}$$

بالقسمة على طرفي (2) على Σx_{2i}^2 نحصل على:

$$\frac{\sum y_i x_{2i}}{\sum x_{2i}^2} = \beta_2 + \beta_3 \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{4}$$

والآن تذكر أن

$$b_{1\,2} = \frac{\sum y_i x_{2i}}{\sum x_{2i}^2}$$

$$b_{3\,2} = \frac{\sum x_{2i} x_{3i}}{\sum x_{2i}^2}$$

المعادلة (4) يمكن كتابتها كالتالي:

$$b_{1\,2} = \beta_2 + \beta_3 b_{3\,2} + \frac{\sum x_{2i}(u_i - \bar{u})}{\sum x_{2i}^2} \tag{5}$$

وبإدخال القيمة المتوقعة على طرفى (5)، نحصل على:

$$E(b_{1\,2}) = \beta_2 + \beta_3 b_{3\,2} \tag{6}$$

حيث تم استخدام التالي: (a) $b_{3\,2}$ تعتبر مقداراً محدداً ثابتًا و (b) β_2، β_3 ثوابت (c) u_i غير مرتبط مع X_{2i} (وغير مرتبط أيضًا مع X_{3i}).

2.A13 توابع استخدام متغير غير مهم: خاصية عدم التحيز:

The Consequences of Including an Irrelevant Variable:
The Unbiasedness Property

وفقًا للنموذج الصحيح (6.3.13)، فإن لدينا التالي:

$$\hat{\beta}_2 = \frac{\sum yx_2}{\sum x_2^2} \tag{1}$$

ونحن نعلم أن المقدار السابق غير متحيز.

للنموذج (7.3.13)، يمكننا أن نحصل على التالي:

$$\hat{\alpha}_2 = \frac{(\sum yx_2)(\sum x_3^2) - (\sum yx_3)(\sum x_2x_3)}{\sum x_2^2 \sum x_3^2 - (\sum x_2x_3)^2} \tag{2}$$

وبالتالي تصبح انحرافات التوزيع الصحيح هي:

$$y_i = \beta_2 x_2 + (u_i - \bar{u})$$

(3) بالتعويض عن Y_i من (3) في (2) وبالتبسيط نحصل على:

$$E(\hat{\alpha}_2) = \beta_2 \frac{\sum x_2^2 \sum x_3^2 - (\sum x_2x_3)^2}{\sum x_2^2 \sum x_3^2 - (\sum x_2x_3)^2} \tag{4}$$

$$= \beta_2$$

أي أن $\hat{\alpha}_2$ مازال غير متحيز.

ويمكننا أن نحصل أيضًا على التالي:

$$\hat{\alpha}_3 = \frac{(\sum yx_3)(\sum x_2^2) - (\sum yx_2)(\sum x_2x_3)}{\sum x_2^2 \sum x_3^2 - (\sum x_2x_3)^2} \tag{5}$$

وبالتعويض عن Y_i من (3) في (5) وبالتبسيط نحصل على:

$$E(\hat{\alpha}_3) = \beta_2 \frac{[(\sum x_2x_3)(\sum x_2^2) - (\sum x_2x_3)(\sum x_2^2)]}{\sum x_2^2 \sum x_3^2 - (\sum x_2x_3)^2} \tag{6}$$

$$= 0$$

وهذه هي قيمتها في النموذج الحقيقي، حيث X_3 غير موجود في النموذج الحقيقي.

3.A13 اثبات المعادلة (10.5.13)

افترض أن لدينا التالي :

$$Y = \alpha + \beta X_i^* + u_i \quad (1)$$

$$X_i = X_i^* + w_i \quad (2)$$

وبالتالي عند إعادة كتابة ذلك في صورة انحرافات نحصل على :

$$y_i = \beta x_i^* + (u_i - \bar{u}) \quad (3)$$

$$x_i = x_i^* + (w_i - \bar{w}) \quad (4)$$

والآن اذا استخدمنا :

$$Y_i = \alpha + \beta X_i + u_i \quad (5)$$

فإننا نحصل على :

$$\hat{\beta} = \frac{\sum yx}{\sum x^2}$$

$$= \frac{\sum[\beta x^* + (u - \bar{u})][x^* + (w - \bar{w})]}{\sum[x^* + (w - \bar{w})]^2}$$

وباستخدام (3) و(4) نحصل على :

$$= \frac{\beta \sum x^{*2} + \beta \sum x^*(w - \bar{w}) + \sum x^*(u - \bar{u}) + \sum(u - \bar{u})(w - \bar{w})}{\sum x^{*2} + 2\sum x^*(w - \bar{w}) + \sum(w - \bar{w})^2}$$

وبما أننا لا نستطيع إدخال القيمة المتوقعة على طرفي المعادلة السابقة، حيث إن توقع خارج قسمة متغيرين لا يساوي قسمه توقعتهما (لاحظ أن: معامل التوقع E هو معامل خطي) فإننا سنقوم أولاً بقسمة كل من البسط والمقام على n ، وتستخدم النهايات (انظر ملحق A لمزيد من التفاصيل عن النهايات). فإن

$$\hat{\beta} = \frac{(1/n)\left[\beta \sum x^{*2} + \beta \sum x^*(w - \bar{w}) + \sum x^*(u - \bar{u}) + \sum(u - \bar{u})(w - \bar{w})\right]}{(1/n)\left[\sum x^{*2} + 2\sum x^*(w - \bar{w}) + \sum(w - \bar{w})^2\right]}$$

الآن ستكون النهاية الاحتمالية لخارج قسمة المتغيرين هي عبارة عن قسمة نهايتهما الاحتمالية. وبتطبيق هذه القاعدة، واستخدام النهاية الاحتمالية لكل مقدار نحصل على التالي :

$$\text{plim}\, \hat{\beta} = \frac{\beta \sigma_{X^*}^2}{\sigma_{X^*}^2 + \sigma_w^2}$$

حيث $\sigma^2_{X^*}$ و σ^2_w هي تباينات X^* و w عندما يزداد حجم العينة، ويصل إلى ما لا نهاية، وقد استخدمنا الحقيقة المثبتة والمتعلقة بأنه عند زيادة حجم العينة، فإنه لا يوجد ارتباط بين u و w كما أنه لا يوجد بينهما وبين قيمة X^* الحقيقية أي ارتباط أيضًا من الشكل السابق نصل إلى:

$$\text{plim}\,\hat{\beta} = \beta\left[\frac{1}{1+(\sigma^2_w/\sigma^2_{X^*})}\right]$$

وهذه هي النتيجة المطلوب إثباتها.

4.A13 اثبات المعادلة (2.6.13)

بما أنه لا يوجد جزء ثابت مقطوع من المحور الصادي، فإنه لتقدير α، وبناء على معادلة الانحدار المار بنقطة الأصل، فإن لدينا التالي:

$$\hat{\alpha} = \frac{\sum X_i Y_i}{\sum X_i^2} \qquad (1)$$

بالتعويض عن Y من النموذج الحقيقي (8.2.13) فإننا نحصل على:

$$\hat{\alpha} = \frac{\sum X_i(\beta X_i u_i)}{\sum X_i^2} = \beta\frac{\sum X_i^2 u_i}{\sum X_i^2} \qquad (2)$$

النظرية الإحصائية تثبت أن $\ln u_i \sim N(0, \sigma^2)$

وبالتالي فإن:

$$u_i = \log\text{ normal }\left[e^{\sigma^2/2},\ e^{\sigma^2}\left(e^{\sigma^2-1}\right)\right] \qquad (3)$$

إذن:

$$E(\hat{\alpha}) = \beta E\left(\frac{\sum X_i^2 u_i}{\sum X_i^2}\right)$$

$$= \beta\left(E\frac{(X_1^2u_1 + X_2^2u_2 + \cdots + X_n^2u_n)}{\sum X_i^2}\right)$$

$$= \beta e^{\sigma^2/2}\left(\frac{\sum X_i^2}{\sum X_i^2}\right) = \beta e^{\sigma^2/2}$$

حيث تم استخدام X's كمقدار غير عشوائي وكل u_i لها قيمة متوقعة تساوي $e^{\sigma^2/2}$

بما أن $E(\hat{\alpha}) \neq \beta$، فإن $\hat{\alpha}$ تعتبر مقدراً متحيزاً للـ β.